Color Tuning in Rhodopsin: A QM/MM Study

Von der Fakultät für Lebenswissenschaften
der Technischen Universität Carolo-Wilhelmina zu Braunschweig
zur Erlangung des Grades
eines Doktors der Naturwissenschaften (Dr.rer.nat.)
genehmigte
D i s s e r t a t i o n

von Jan Stephan Frähmcke
aus Itzehoe

1. Referent: Professor Dr. Peter Jomo Walla
2. Referent: Professor Dr. Marcus Elstner
eingereicht am: 21.12.2011
mündliche Prüfung (Disputation) am: 03.05.2012

Druckjahr 2012

Vorveröffentlichungen der Dissertation

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakultät für Lebenswissenschaften, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

Publikationen

WANKO, M., HOFFMANN, M., FRÄHMCKE, J., FRAUENHEIM, T., ELSTNER, M., **2008**. *Effect of polarization on the opsin shift in rhodopsins. 2. Empirical polarization models for proteins*, J. Phys. Chem. B, 112(37), 11468-11478.

FRÄHMCKE, J.S., WANKO, M., PHATAK, P., MROGINSKI, M.A., ELSTNER, M., **2010**, *The Protonation State of Glu181 in Rhodopsin Revisited: Interpretation of Experimental Data on the Basis of QM/MM Calculations*, J. Phys. Chem. B, 114(34), 11338–11352.

FRÄHMCKE, J.S., WANKO, M., ELSTNER, M., **2012**, *Building a Model of the Blue Cone Pigment Based on the WT Rhodopsin Structure with QM/MM Methods*, J. Phys. Chem. B, 116(10), 3313–3321.

Tagungsbeiträge

Is Glutamate^{181} *Protonated in Bovine Rhodopsin's Dark State? A QM/MM Study*
Jan S. Frähmcke, Marius Wanko, Michael Hoffmann, Thomas Frauenheim, Marcus Elstner, Poster at WORKSHOP ON COMPUTER SIMULATION AND THEORY OF MACROMOLECULES **2008**, Hünfeld, Germany.

Spectral Tuning in Rhodopsin and Cone Visual Pigments Calculated with QM/MM-Methods
Talk at WORKSHOP ON COMPUTER SIMULATION AND THEORY OF MACROMOLECULES **2008**, Hünfeld, Germany.

Is Glutamate^{181} *Protonated in Bovine Rhodopsin's Dark State? A QM/MM Study*
Jan S. Frähmcke, Marius Wanko, Michael Hoffmann, Thomas Frauenheim, Marcus Elstner, Poster at 44TH SYMPOSIUM ON THEORETICAL CHEMISTRY STC **2008**, Ramsau am Dachstein, Austria.

Is Glutamate^{181} *Protonated in Bovine Rhodopsin's Dark State? A QM/MM Study*
Jan S. Frähmcke, Marius Wanko, Michael Hoffmann, Thomas Frauenheim, Marcus Elstner, Poster at WORKSHOP ON COMPUTER SIMULATION AND THEORY OF MACROMOLECULES **2009**, Hünfeld, Germany.

Spectral Tuning in Rhodopsin and Cone Visual Pigments Calculated with QM/MM-Methods
Jan S. Frähmcke, Michael Hoffmann, Marius Wanko, Thomas Frauenheim, Marcus Elstner, Poster at WORKSHOP ON COMPUTER SIMULATION AND THEORY OF MACROMOLECULES **2010**, Hünfeld, Germany.

Color Tuning in Blue Cone Visual Pigment. A QM/MM Study
Jan S. Frähmcke, Marcus Elstner, Poster at 2ND CHANNELRHODOPSIN AND LIGHT-GATED ENZYMES CONFERENCE **2010**, Hiddensee, Germany.

Contents

List of Figures

List of Tables

Abbreviations

Shortcut	Legend
AA	amino acid
AANO	approximate average natural orbitals
act	active MO in DDCI
AO	atomic orbital
BB	backbone
BLA	bond-length alternation
CASSCF	complete-active-space self consisting field method
CASPT2	CASSCF method with perturbation theory
C-I – III	cytoplasmic loops I – III
CCI	complex counterion
CHARMM	Chemistry at HARvard Macromolecular Mechanics (force field and program package)
CI	counterion
CI	configuration interaction
CID	configuration interaction with double excitations
CIS	configuration interaction with single excitations
CISD	configuration interaction with single and double excitations
CSF	configuration state function
CT	charge transfer
DFT	density functional theory
DFTB	density functional based tight binding
E-I – III	extracellular loops I – III
EE	electrostatic embedding
ext	external MO in DDCI
exp.	experimental
FTIR	fourier transformed IR
GROMOS	GROningen MOlecular Simulation (force field)
HBN	hydrogen bonded network
HF	hartree Fock theory
HOMO	highest occupied molecular orbital
int	internal MO in DDCI
IR	infra-red light
L	large
LH I, LH II	bacterial light-harvesting complexes
LHC I, LHC II	light-harvesting complexes in plants
LUMO	lowest unoccupied molecular orbital
M	medium sized
man	SORCI calculation with manual selection of π orbitals
MD	molecular dynamics
MM	molecular mechanics
MO	molecular orbital
MP2	Møller-Plesset perturbation theory in second order

Shortcut	Legend
nat3	SORCI calculation with CAS(2,2) in an MRDDCI2 precalculation
nat19	SORCI calculation with CAS(4,4) in an MRDDCI3 precalculation
OM2	Orthoganization Method 2[1]
opt	optimized
orca	program package used for SORCI calculations[2]
PB	Poisson-Boltzmann
pert	perturbation
PS	photosystem in plants
QM	quantum mechanics
QMx	QM region
RAS	restricted active space (e.g. (12: 5 2/2/5 2)
RC	reaction center in photosystem II
Ret	retinal (π-system)
Rh	rhodopsin
S	small
S_0	electronic ground state
$S_{1,2,...}$	electronic excited state
SC	sidechain
SCF	self-consistent-field
SE	Schrödinger equation
SORCI	Spectroscopy Oriented Configuration Interaction[3]
TM	transmembranal
UV	ultra-violet light
UV–Vis	ultra-violet and visible light
vert	vertical
Wat	water
WT-Rh	wild-type Rhodopsin
XL	very large
XS	very small

Constants

c	$= 2.998 \times 10^{8}\,\mathrm{m{\cdot}s^{-1}}$	Speed of light
e	$= 1.602 \times 10^{-19}\,\mathrm{C}$	Elementary charge
ε_0	$= 8.854 \times 10^{-12}\,\mathrm{C^2{\cdot}J^{-1}{\cdot}m^{-1}}$	Dielectric constant of vacuum
h	$= 6.626 \times 10^{-34}\,\mathrm{J{\cdot}s}$	Planck constant ($\hbar = \frac{h}{2\pi} = 1.055 \times 10^{-34}\,\mathrm{J{\cdot}s}$)
k	$= 1.381 \times 10^{-23}\,\mathrm{J{\cdot}K^{-1}}$	Boltzmann constant
m_e	$= 9.109 \times 10^{-31}\,\mathrm{kg}$	Elektron mass
m_p	$= 1.673 \times 10^{-27}\,\mathrm{kg}$	Proton mass

Units

symbol	meaning	unit
B	crystallographic temperature factor	[Å^2]
m	Mass	[kg]
$\vec{\mu}$	Dipole moment	[D]
μ_x	Dipole moment in x-direction	[D]
$\Delta\vec{\mu}$	Difference dipole moment	[D]
t	Time	[s]
ε	Dielectric constant	[C^2·J^{-1}·m^{-1}]
q	Charge	[$e = 1.602 \times 10^{-19}$ C]
p	Pressure	[Pa = N·m^{-2}]
V	Volume	[L]
T	Temperature	[K]
N	Amount of substance	[mol]

Abbreviation and Properties of Amino Acids.

Name	Code	properties
Alanine	Ala, A	nonpolar, XS
Arginine	Arg, R	polar, L, titratable
Asparagine	Asn, N	polar, M
Aspartic acid	Asp, D	polar, M, titratable
Cysteine	Cys, C	polar, S
Glutamic acid	Glu, E	polar, L, titratable
Glutamine	Gln, Q	polar, L
Glycine	Gly, G	nonpolar, XS
Histidine	His, H	polar, L, titratable
Isoleucine	Ile, I	nonpolar, L
Leucine	Leu, L	nonpolar, L
Lysine	Lys, K	polar, XL, titratable
Methionine	Met, M	nonpolar, L
Phenylalanine	Phe, F	nonpolar, L
Proline	Pro, P	nonpolar, M
Serine	Ser, S	polar, S
Threonine	Thr, T	polar, M
Tryptophan	Trp, W	nonpolar, XL
Tyrosine	Tyr, Y	polar, XL
Valine	Val, V	nonpolar, M

Chapter 1

Introduction

1.1 Proteins Using Light

In biologigal systems proteins play a key role and many of them have special tasks as catalysis, structural or mechanical functions, cell signaling or ion pumping. These reactions need energy that is produced by cellular respiration in animals and by photosynthesis in plants, where proteins produce energy from light. Also some bacteria use sunlight as an energy source. All these processes are common in producing ATP (adenosine triphosphate), the "molecular unit of currency" in all living organisms. However, light is not only used for energy production but also for sensoric purposes. Animals own photoreceptors in their eyes that detect light and give signals to the brain. There are bacteria and algae using photoreceptors to find a place with good light intensity. This process is called phototaxis. These small organisms live in water, "swim" around and stop swimming if they find good light conditions (positive phototaxis) or they "swim" away if conditions are bad, e.g. too high light intensity (negative phototaxis). This movement is not a directed movement to or away from light but a random search for good locations.

There are many proteins using light to produce energy. E.g. photosynthetic light-harvesting complexes as the bacterial light-harvesting complexes I and II (LH I, LH II) and photosystems from plants (see PS II in Figure 1) that are built of light-harvesting complexes I and II (LHC I, LHC II) and the reaction center (RC). In all these proteins the chromophores are carotenoids and (bacterio-)chlorophylls that are non-kovalently bound to the protein.

Additionally, energy is produced from sunlight by bacteriorhodopsin (bR) in some bacteria. It is a protein that absorbs light and pumps a proton through the membrane in which it is located. The produced proton gradient starts ATP-synthase to produce ATP.

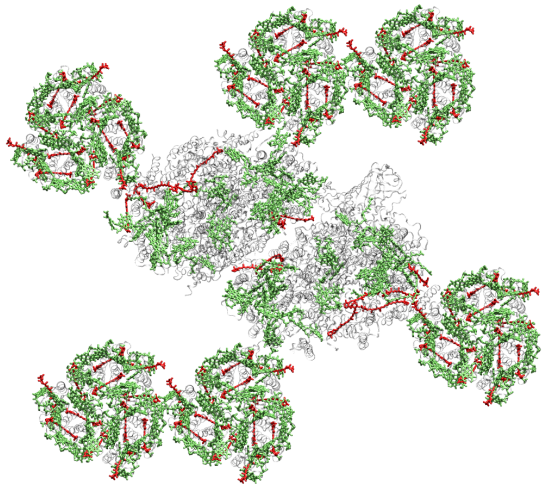


Figure 1: Constitution of plant's photosystem II (PS II, pdb:1S5L[4]). Antenna complexes formed by trimers of the light-harvesting complex II (LHC II) are locatd around the reaction center (RC). The RC is built of two subunits with seven carotenoids (red) and 35 chlorophylls (green). An LHC II contains four carotenoids and 14 chlorophyll molecules.

1.2 Color Tuning

In both examples (PS II and bR) the adaption what kind of light is absorbed is important. Primarily the available light should be used efficiently and secondly the light should have enough energy to start the complex cascade of reactions ending in the production of ATP. Light with too less energy (infra-red, IR) is not able to induce the structural changes needed for producing ATP and light with too much energy (ultra-violet, UV) can produce dangerous side products that can harm the bacteria or plants. Light-harvesting complexes absorb blue and red light (Soret- and Q-band of chlorophylls) and green-blue light (S_2 absorption of carotenoids). Bacteriorhodopsin absorbs green light at about 570 nm.[5]

The modulation of the wavelength of the absorbed light is called color tuning. It is very important for color vision (described in Chapter 1.3). Two or more types of photoreceptor proteins absorb different colors although they are very similar in structure and function. The reasons and mechanisms why different colors are absorbed are investigated in this work.

Color tuning can work in two different ways. First, a different chromophore can absorb different light. This mechanism of color tuning can not be used extensively because the chromophore also has other functions, e.g., conformational changes, producing or quenching other species. Thus, the chromophore can not be exchanged by another one because of sterical reasons, too. Proteins are perfectly designed to hold a special chromophore.

The second method of color tuning works with the same chromophore and only the surrounding protein, the so-called binding pocket, is changed. This method of color tuning is the main topic of this work. The protein changes result in electrostatic and steric effects.

1.3 Vision and Retinal Proteins

Vision of vertebrates takes place in the eyes. The incoming light is focused by adjustable lenses to the retina which contains two kinds of photoreceptor cells, rods and cones. The light is transformed into electro-chemical impulses being transmitted via the optic nerve to the visual cortex in the brain. The photoreceptor cells were first distinguished from each other on the basis of the morphological shape of their outer segment regions. Vertebrates have two, three, or sometimes four types of cone cells. Having more than one type of cone cells allows for distinguishing different colors.There are three types of cones in the human retina, each of which has a visual pigment with a distinctive absorption maximum (red, green, and blue). Cone pigments are less sensitive than rod cells and work under daylight conditions. There is only one type of rod cells. These are more sensible and more numerous. Thus, in dim light, colors can not be distinguished any more.[6]

The cone and rod receptor cells are similar (see Figure 2). They consist of the outer membrane region, where in the case of rod cells disks build from membranes are inside the cell and in the case of cone cells invaginations form membrane disks. A protein called opsin is embedded in these membranes. Opsin together with the pigment molecule retinal forms the visual pigment called rhodopsin (Rh) and iodopsin for rod and cone cells, respectively.

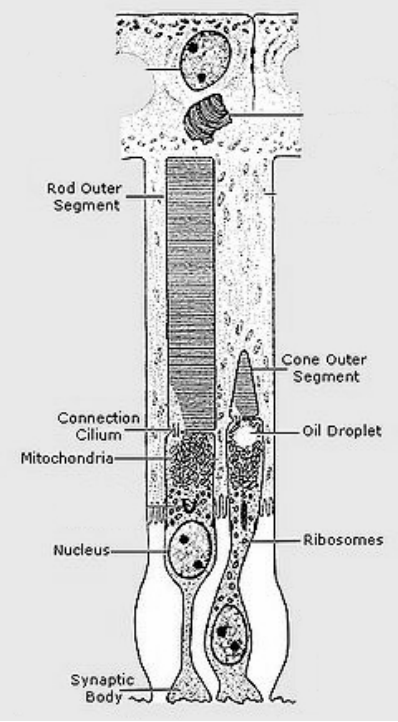


Figure 2: Functional parts of the photoreceptor cells (changed from Young[7])

The opsin proteins of rhodopsin and the iodopsins belong to a superfamily of seven-transmembrane (TM)-helix integral membrane proteins known as the G protein-coupled receptors (GPCRs). GPCRs are very common and are not only responsible for vision, but also for tasting and smelling. Upon activation by a

ligand, they bind G proteins that initiate intracellular signaling cascades. Activation of G proteins by GPCRs requires the propagation of structural signals from the receptor-binding domain to the G protein. In the vertebrate visual system, the ligand is the chromophore 11-*cis* retinal. It is covalently bound to the opsin protein as a protonated Schiff base (PSB) via the ϵ-amino group of a lysine (Lys, K) residue in helix VII (see Figure 3). A counterion is located close to the PSB that consist of Glu113,[8,9] a water molecule and some other residues.

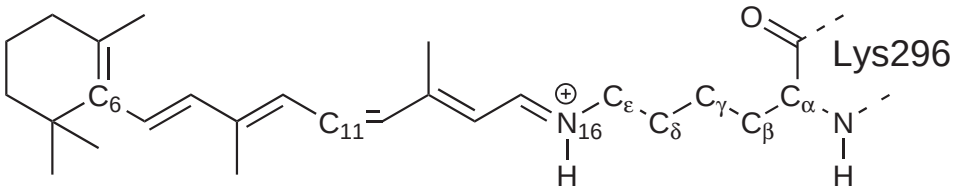


Figure 3: Structure and nomenclature of important atoms in retinal building a protonated Schiff base (PSB) with Lys296.

Biochemical and biophysical analysis of rhodopsin[10] have provided a wealth of information on various aspects of its structure and mechanism of signaling, such an understanding of the human cone pigments is missing because these pigments are much more seldom in nature.

The mechanism of signal conversion after absorption of light is not a main topic in this work but is described briefly to understand some experimental results being important for other topics (detailed description in a review by Birge[5] and references therein):
The 11-*cis*-retinal absorbs light and is then changing the configuration of the C_{11}–C_{12} double bond to *trans*. Due to this change, conformational changes in the protein binding site are induced that produce a series of spectroscopically distinguishable states with different life times: Photorhodopsin (ps), Bathorhodopsin (ns), Lumirhodopsin (μs), Metarhodopsin I (Meta I, ms), Metarhodopsin II (Meta II, s). The Meta I and Meta II states are in equilibrium with each other and only the Meta II state is the only state that could activate the heterotrimeric G protein, transducin. One rhodopsin molecule in the Meta II state can activate up to 1000 transducin molecules. These block cation-specific ion channels and produce a hyperpolarization of the plasma membrane of the rod or cone cells, resulting in a more negative potential inside the cell. This potential difference starts the nerve impulse. Deactivation of the Meta II state includes the expulsion of the isomerized all-*trans*-retinal from the binding site. An enzyme is needed to convert the all-*trans*-retinal to 11-*cis*-retinal again that can be rebound to the opsin protein to rebuild the ground state rhodopsin again. In the transition from Meta I to Meta II, the PSB is deprotonated. The proton from the PSB is transfered to the counterion (Glu113) resulting in strongly changed conformation because of a much weaker interaction of the Schiff base with the protonated (neutral) counterion. The structure of the Meta I and II states as well as the transition mechanism are unknown. They were investigated spectroscopically with UV–Vis, Raman[11] and FTIR spectroscopy[12] leading to different results. These are in

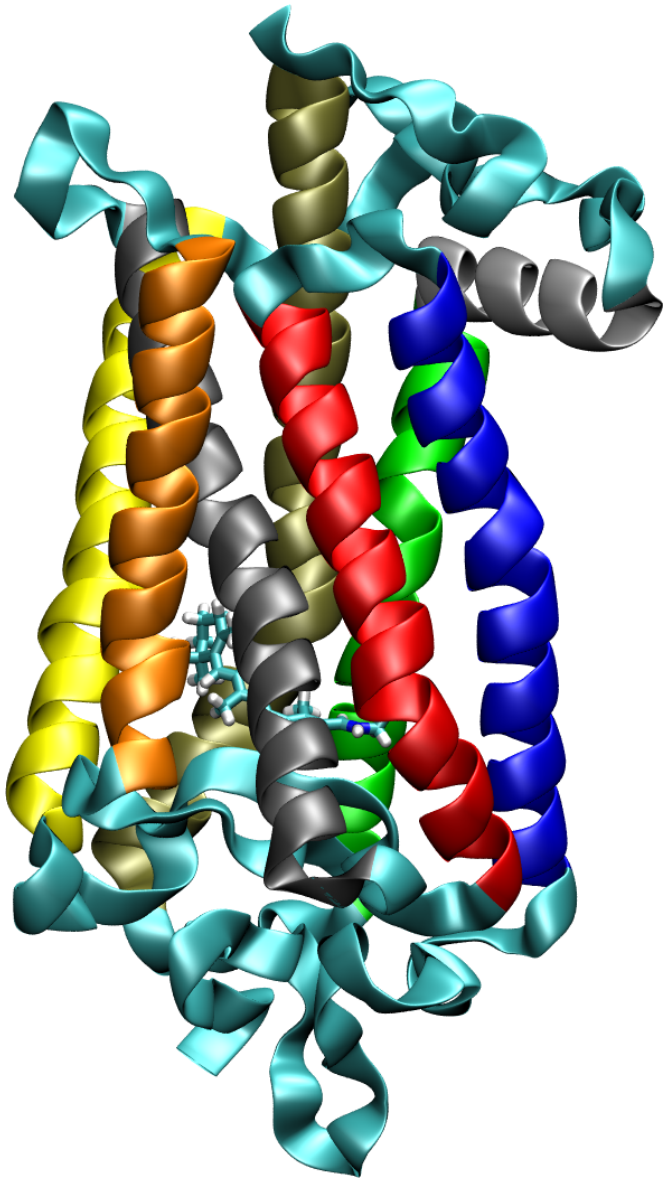
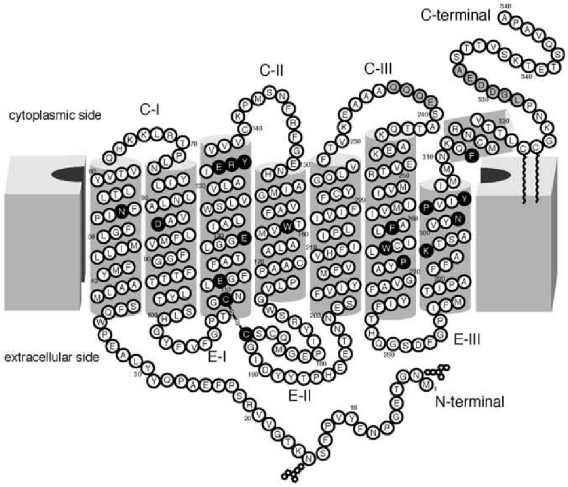


Figure 4: Model of Rh with transmembranal helices marked in colors and the retinal chromophore located in the middle of the protein.

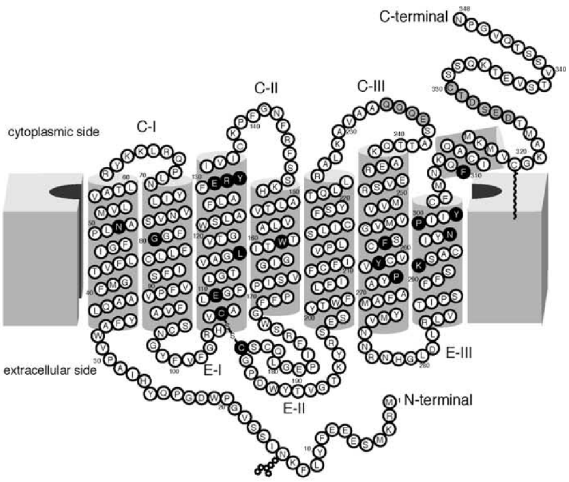
detail described in Section 4.1.2.

The opsin molecules in human rhodopsin and the three types of iodopsins are very similar. All have seven transmembranal helices (I – VII), and three extracellular and cytoplasmic loops (E-I – E-III and C-I – C-III) giving a total of about 350 aminoacids (see Figure 4). The secondary structure of rhodopsin and the three iodopsins was first drawn twodimensionally by Kosower in 1988.[13] With new informations from crystal structures of rhodopsin[14–16] Stenkamp et al.[17] drew new pictures (Figure 5). Blue, green, and red iodopsins are homologous to rhodopsin by 41, 38, and 37%. Green and red iodopsin are very similar and differ only in 15 aminoacids. The homology of $< 50\%$ seems to be a small homology but many of the differing aminoacids are still similar. An alignment of the primary structures of Rh and the cone opsins is possible without problems (see Figure 5). The numbering of the aminoacids is different in Figure 5 because of different lengths of the C- and N-termini. For more clarity, in this work the numbering of rhodopsin

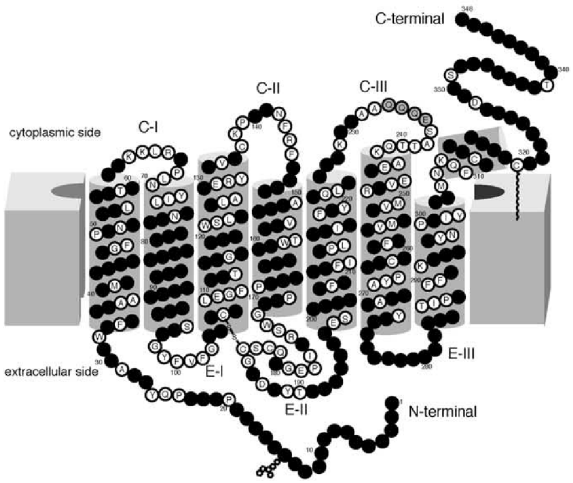
is used for the iodopsins, too.



(a) Rhodopsin, key residues are marked in black



(b) Blue iodopsin



(c) Blue iodopsin vs. rhodopsin, differing aminoacids are marked in black

Figure 5: Two-dimensional models of rhodopsin (a), blue iodopsin (b), and the comparison of rhodopsin and blue iodopsin (c). Pictures taken from Stenkamp et al.[17]

The differences in protein structure tune the color of light that is absorbed by the chromophore in the protein. Rhodopsin absorbs at 498 nm (2.49 eV),[18] the three cone pigments absorb at 560 nm (2.21 eV, human red cone (HR)),[19] 530 nm (2.33 eV, human green cone (HR)),[19] and 414 nm (2.99 eV, human blue cone (HB)).[20]

The mechanism of color tuning is complicated. Electrostatic effects play a key role, but also steric effects changing the geometry of the chromophore can have an influence on the light absorbance. Changes in the residues surrounding the

counterion (Glu113) can also have a large effect electrostatically or directly by changing the electronic structure of the retinal chromophore. The influence of the calculation method and electrostatics on the chromophore is studied in Chapter 3 and the investigation on the color tuning mechanisms between Rh and HB is the topic of Chapter 5.

The residue 181 is a very important residue. It is a glutamic acid (Glu, E) in rhodopsin and blue iodopsin, and a histidine (His, H) in red and green iodopsins. It is supposed that the positive charge of the histidine is neutralised by a chloride ion (Cl^-). The exact position of this Cl^- is not known.[21] This work concentrates on rhodopsin and blue iodopsin because of this problem.

The protonation state of Glu181 is disputed. Chapter 4 studies this topic and in the next section, the topic of protonation states of amino acids in proteins is viewed in detail.

1.4 Proteins

Proteins are macromolecules built of amino acids (AAs) connected via peptide bonds with an amide group (–C_αHR–C(=O)–NH–C_αHR–). The term amino acid is a very general one. It states that a molecule contains an amino group (NH_2) and a carboxylic acid (COOH). Proteins are built of α-amino acids (mainly in L-configuration) that can be abbreviated by the formula H2N–C_αHR–COOH. R represents a sidechain. 22 proteinogenic amino acids are known (see Table for a list and their abbreviations). They can be divided in classes by several attributes like size, polarity, aromaticity, and charge. Free amino acids with neutral side chains are present in a zwitterion form at neutral pH: The amino group is protonated and the carboxylic acid is deprotonated. In proteins the protonation state of these groups is not changeable, because both of them are connected via an amide bond.

1.4.1 Protonation State of Amino Acids

The protonation state of amino acids is a major topic in this work. It is only about the protonation state of amino acids that have a chargable (titratable) sidechain (Arg, Asp, Glu, His, and Lys, see Figure 6). The protonation states of titratable amino acids in rhodopsin are discussed in Section 1.4.3. The special case of Glu181 is discussed in Chapter 4. It is viewed from the experimental and theoretical perspective giving the main conclusion of this chapter that Glu181 is deprotonated (charged) in Rh.

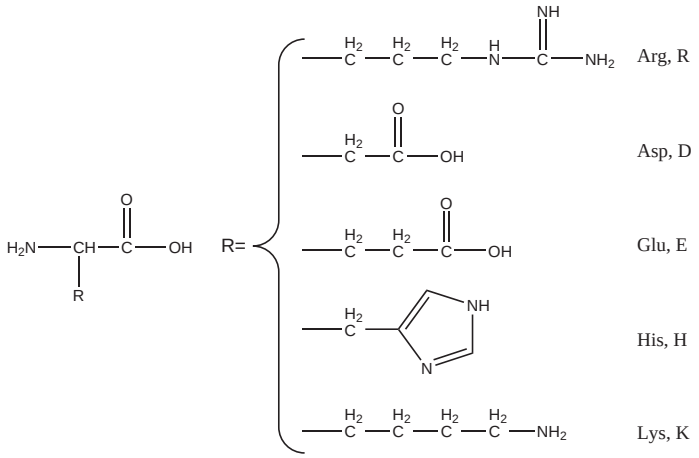


Figure 6: Titratable amino acids (arginine, aspartic acid, glutamic acid, histidine, lysine).

The protonation state of an amino acid sidechain is not an absolute property. It is a counterplaying process and the relative speed of protonation and deprotonation gives the equilibrium amount of protonated and deprotonated molecules. For a single molecule only a probability can be given.

The group that makes the sidechain titratable is either a carboxylic acid (–COOH) for glutamic acid (Glu, E) and aspartic acid (Asp, D), a guanidinium group (–NH–C(=NH)–NH_2) for arginine (Arg, R), an imidazole group for histidine (His, H), or an amino group (–NH_2) for lysine (Lys, K). Table 1 gives the pK_a values and standard protonation states of these five AAs at pH7.

Table 1: pK_a values,[22] protonation states and charges of the five titratable amino acids at pH7 (prot.: protonated, deprot.: deprotonated, unprot.: unprotonated).

	Asp	Glu	His	Lys	Arg
pK_a	4.1	4.1	6.0	10.8	12.5
protonation state	deprot.	deprot.	unprot.	prot.	prot.
charge	negative	negative	neutral	positive	positive

The protonation state of titratable AAs in proteins is not only depending on the pH of the solution but also on the environment. Proteins in aquous solution most probably have all titratable AAs in their standard protonation state because they are in direct contact to water. The sidechains of small or medium sized proteins that have no tertiary structure are also fully accessible to water and their protonation states completely depend on the pH of the solution. Larger proteins have a tertiary structure and build a hydrophobic core not being accessible to water. Therefore, the protonation state of AAs is partly independent from the pH of the solution. Instead, the protonation state is determined by the surrounding protein. The polarity and protonation states of other AAs and embedded water molecules can change the protonation states. Titratable AAs in an apolar surrounding (without water and polar AAs) tend to be in their uncharged pro-

tonation state. Pairs of titratable AAs can stabilize their charged forms together. These salt bridges of negatively and positively charged AAs increase the protein stability and are important for the tertiary and also quaternary structure.

X-ray structures of proteins do not include the small hydrogen atoms and therefore the protonation states of titratable AAs have to be determined with other methods to perform theoretical studies with a threedimensional model of a protein. When using an X-ray structure one always has to keep in mind that also the determination of the X-ray structure is a kind of calculation, fitting the atoms into the measured electron density. The method of modeling titratable AAs is often not published. But the choice of method can have a large impact on the predicted structure. And also the predicted structure can have a large impact on the protonation state.

The protonation states of titratable AAs can be determined experimentally and theoretically as described in the next section.

1.4.2 Determination of Protonation State of Amino Acids

IR spectroscopy can be used to determine whether an AA is charged or uncharged. But this can not be done directly because there are many C=O and N–H bonds in the protein backbone and thus the bands of charged or uncharged carboxylic acids or guanidinium, imidazole, and amine groups can not be observed in an IR spectrum. But a determination is possible indirectly due to missing bands when mutating the respective AA. In the following the determination will be described for aspartic and glutamic acids. For other titratable AAs the procedure is similar but different bands have to be viewed.

Aspartic and glutamic acid produce C=O stretch bands clearly above $1700\,\mathrm{cm}^{-1}$ ($1712 - 1788\,\mathrm{cm}^{-1}$, depending on the hydrogen bonding situation) if they are protonated but the deprotonated AAs give two bands at ≈ 1400 and $\approx 1570\,\mathrm{cm}^{-1}$.[23] Mutating an aspartic or glutamic acid into an asparagin (Asn, N) or glutamine (Gln, Q) gives a new band at $\approx 1680\,\mathrm{cm}^{-1}$ from the C=O bond in the amide group.[23] But this band is clearly overlapping with the strong amide I band produced by all C=O groups in the backbone amide bonds. For the determination if an aspartic or glutamic acid is deprotonated or not the missing band in the spectrum of the D→N or E→Q mutant is decisive.

Even more indirectly one can measure other spectroscopic properties (Raman, UV-Vis, NMR, etc.) and interpret the differences between the wild-type (WT) protein and the mutant. Mostly the spectroscopic properties of a protonated aspartic or glutamic acid are very similar to the amide analog, because both AAs are uncharged and have similar hydrogen-bonding potential. The difference between a negatively charged AA and the uncharged amide analog are expected to be very large. This explanation has to be used very carefully. Sometimes the effect of the difference in charge can be surprisingly small and the other way round the expected small difference in the hydrogen-bonding properties of the

protonated acid and the amide can have a large impact on spectroscopic properties. Thus the choice of the protonation state can be wrong and therefore a combination of methods should be used.

Computationally the protonation state of titratable AAs can be determined directly via calculation of the pK_a in their special surrounding. Therefore, the free energy difference of the protonated and unprotonated form has to be calculated via MD simulations. Due to the large difference between COOH and COO^- (charge, hydrogen bonding capabilities, etc.) the MD simulations have to be very long. The negative charge or the changed hydrogen-bonding pattern without the hydrogen atom can induce large structural changes in the surroundings and thus the free energy difference is very hard to calculate.

Another way to computationally determine the protonation state of a titratable AA is to calculate spectroscopic properties that have been or will be measured experimentally. This can help to interpret and show errors in the interpretation of the experimental results.

1.4.3 Protonation State of Amino Acids in Rhodopsin

In the primary sequence of Rh there are many titratable AAs. Most of them are located in the extracellular or cytoplasmic loops. These loops are located in regions where bulk water is present and thus, titratable AAs are in their standard protonation state (see Table 1).

Table 2: Location of titratable AAs in WT-Rh in the bulk water regions.

N-Terminus	Loops						C-Terminus
	C1	E1	C2[a]	E2	C3[b]	E3	
Glu5	His65	His100	Lys141	Arg177	Lys231	His278	Lys311
Lys16	Lys66		Arg147	**Glu181**	Glu232	Asp282	Arg314
Arg21	Lys67		Glu150	Asp190	Glu239		Lys325
Glu25	Arg69			His195			Asp330
Glu33				Glu196			Asp331
				Glu197			Glu332
				Glu201			Lys339
							Glu341

[a] Glu134 and Arg135 are located in helix3 and His152 in helix 4 but located in the bulk water region close to loop C2.
[b] Lys245, Glu247, Lys248, Glu249, and Arg252 are located in helix6 but located in the bulk water region close to loop C3.

In the helical region there are only few titratable AAs. Asp83 and His211 are in the central regions of helix2 and helix5, respectively. His211 is modeled in its standard protonation state (unprotonated, neutral). The question of Asp83 is a bit more complicated. The environment is very polar (some water molecules

and polar AAs) but there are no potential positive charges close by. Thus, Asp83 is modeled protonated, neutral. The last two titratable residues are Glu113 and Lys296. Lys296 is the special residue, the retinal molecule is covalently bound to. It is known to form a protonated Schiff base (PSB). Thus this residue is positively charged. Glu113 is known to be the primary counterion of the PSB and is deprotonated, (negatively charged).

Glu181 is given in bold face in Table 2 because it is a special case. Chapter 4 will be concerned with this AA in detail.

Counting together all charged AAs, the total charge is not zero but $-2\,\mathrm{e}$. And also the distribution of the charges is not equal but a net charge of $-6\,\mathrm{e}$ and $+4\,\mathrm{e}$ on the cytoplasmic and extracellular side, respectively. But this is no reason to change the protonation state of some AAs because in bulk water there are many cations and anions that can counterbalance these net charges.

Chapter 2

Computational Methods

Several computational methods were used in this work. Many of them are standard methods in theoretical chemistry and are only briefly described (Molecular Mechanics (Section 2.1.1), Hartree-Fock (Section 2.1.3.1), DFTB (Section 2.2.2), molecular dynamics simulations (Section 2.2.1.1), QM/MM (Section 2.1.2). Also methods describing calculation of excitation energies are mentioned, that are developed by other groups and are described in detail in the given literature (OM2/MRCI (Section 2.2.3.2), SORCI (Section 2.2.3.3)).

2.1 Basics on Computational Chemistry

The task of computational chemistry is to calculate special properties of molecules as acurate as possible. The physics behind molecules is very complicated and very long calculation times are needed. Therefore, methods were developed to describe the physics slighly less acurate but much faster. A general classification can be quantum-mechanical methods that include electrons explicitly and molecular mechanics methods.

2.1.1 Molecular Mechanics

In molecular mechanics methods (MM) electrons in molecules or atoms are not considered explicitly. Atoms are connected with each other (connectivity) and bond lengths, bonding angles and dihedral angles determine the molecular geometry. With these geometrical parameters the molecular energy is determined. Additionally, the total energy is calculated with non-bonding interactions of atoms that are not connected (e.g. different parts of a molecule come close to each other).

Equations and parameters are needed to calculate the molecule's total energy. The combination of equations and parameters is called force field. There are

different force fields available to be used for different purposes. For proteins and macromolecules the GROMOS (GROningen MOlecular Simulation) and CHARMM (Chemistry at HARvard Macromolecular Mechanics) force fields are the best known and in this work the CHARMM force field is used because of numerous benchmark works of CHARMM in combination with DFTB. These works will be given in the following section about QM/MM with DFTB/CHARMM (Section 2.1.2).

With typical force fields the molecule's total energy is calculated as

$$\begin{aligned} E^{\text{total}} =& E^{\text{bonded}} + E^{\text{nonbonded}} \\ =& \sum_{\text{bonds}} \frac{1}{2} k_{\text{b}} (r - r_0)^2 \\ &+ \sum_{\text{angles}} \frac{1}{2} k_{\text{a}} (\theta - \theta_0)^2 \\ &+ \sum_{\text{torsions}} \frac{1}{2} V_n \left[1 + \cos(n\omega - \omega_0)\right] \\ &+ \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} \epsilon_{ij} \left[\left(\frac{r_{0ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{0ij}}{r_{ij}} \right)^{6} \right] \\ &+ \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \end{aligned} \tag{1}$$

The first three terms (bonding terms) are for bond lengths, bonding angles, and torsional angles. The latter two terms are the nonbonded terms and are for van der Waals and electrostatic interactions of all atoms that are not bonded to each other.

The parameters needed for calculating the total energy are k_{b} (bond length force constant), r_0 (equilibrium bond length), k_{a} (bonding angle force constant), θ_0 (equilibrium bonding angle), n, V_n, and ω_0 determining the number, depth, and location of minima for torsional barriers, ϵ_{ij} and r_{0ij} describing the depth and location of the minimum in the Lennard-Jones potential, and q_i (partial atomic charges). The q_i are atomic properties. All other parameters are defined pairwise, triplewise or quadruplewise depending on the atom types being involved. For all parameters the kind of bonding situation is important. E.g., a C–C single bond and a C=C double bond have different r_0 and especially the number of minima of the torsional energy around this bond is different ($n = 3$ and 2, respectively). Thus, for every element (H, C, N, ...) several atom types (e.g. H, HA, HB, HP for polar, nonpolar, backbone, aromatic hydrogen atoms) are available and depending on the connected atom versions, different parameters are used.

The nonbonded interaction would make the calculation of large molecules slow due to the large number of atom pairs and thus a cut-off distance for the last two terms is very often used.

Having the coordinates of all atoms it is computationally simple to calculate all terms and get the total energy. Also gradients are easy to calculate and thus geometry optimizations and molecular dynamics simulations (Section 2.2.1.1) are a standard feature of molecular mechanics calculations.

For small molecules it is a practicable way to define atom versions for every atom and then the program takes the coordinates and parameters and calculates the total energy. But for larger molecules as proteins the number of atoms is very large and the work to manually select atom versions would take very long. Additionally the atomic charges are not fixed for every atom version and has to be assigned also manually. A complex program package is needed to do all this work automatically for large molecules.

CHARMM is not only the name for the force field but also for the program package to do the calculation and setup tasks. Assigning atom versions and atomic charges to all atoms in a protein is a standard task of CHARMM. It is done via a topology file where informations for atom version and atomic charges for all atoms in a protein are stored. For every amino acid the names of the atoms are defined and these names are also given in crystal-structure files from the pdb-database. The topology file has a list of all amino acids and therein a list of all atom names. The atom versions and atomic charges are in the third and fourth column. The atoms are merged to groups whose charge is exactly zero or exactly ± 1 for charged amino acids.

```
GROUP                                                  GROUP
ATOM N    NH1    -0.47  !     |                        ATOM N    NH1    -0.47  !     |
ATOM HN   H       0.31  !  HN-N                       ATOM HN   H       0.31  !  HN-N
ATOM CA   CT1     0.07  !     |     HB1               ATOM CA   CT1     0.07  !     |   HB1
ATOM HA   HB      0.09  !     |    /                  ATOM HA   HB      0.09  !     |   |
GROUP                   !  HA-CA--CB-HB2              GROUP                   !  HA-CA--CB--OG
ATOM CB   CT3    -0.27  !     |    \                  ATOM CB   CT2     0.05  !     |   |   |
ATOM HB1  HA      0.09  !     |     HB3               ATOM HB1  HA      0.09  !     |   HB2 HG1
ATOM HB2  HA      0.09  !   O=C                       ATOM HB2  HA      0.09  !   O=C
ATOM HB3  HA      0.09  !     |                       ATOM OG   OH1    -0.66  !     |
GROUP                   !                           ATOM HG1  H       0.43
ATOM C    C       0.51                              GROUP
ATOM O    O      -0.51                              ATOM C    C       0.51
                                                     ATOM O    O      -0.51
```

MM methods are very reliable for large biomolecules with standard residues.[24] But there are some problems with MM methods. Reactions or even changes in bonding situations (proton transfer) cannot be calculated with MM methods because of atom types change during reactions. A second problem is that good parameters for all parts of the protein are needed for a calculation. For retinal proteins this is problematic because retinal is such an unusual molecule that it is not included in the usual topology file of CHARMM. To use it, the assignment of the atom types and especially atomic charges are needed. But the charges stongly depend on the geometry and the electrostatic environment. There are a topology file and parameters available[25] but these should only be used if the main topic of research is not the retinal geometry or for long MD simulations.

In cases where the retinal geometry is important a more sophisticated method should be used. But this is not possible for the whole molecule as it is limited

to current computational speed. The development of multi-scale methods has solved this problem. A small, important part of the big molecule is calculated with a high-level method (quantum mechanics) and only the remaining protein is calculated with the less acurate but very fast MM method.

2.1.2 QM/MM

There are two concept in combining quantum mechanics (QM) and MM. The first is a subtractive QM/MM coupling (e.g. Morukuma's ONIOM scheme[26]). Therein the whole system is calculated with MM, the interesting region (called QM region) is calculated both with MM and QM. The total energy is then calculated via

$$E_{\text{all}}^{\text{QM/MM}} = E_{\text{all}}^{\text{MM}} + E_{\text{QM}}^{\text{QM}} - E_{\text{QM}}^{\text{MM}} \tag{2}$$

where the upper index denotes the calculation method and the lower index the region (QM, all).

The second scheme which is used in this work is the additive scheme. The QM region is deleted in the MM calculation and instead calculated with the QM method. The total energy is then given by

$$E_{\text{all}}^{\text{QM/MM}} = E_{\text{MM}}^{\text{MM}} + E_{\text{QM}}^{\text{QM}} + E_{\text{QM–MM, vdW}}^{\text{MM}} + E_{\text{QM–MM, ES}}^{\text{MM or QM}} \tag{3}$$

where the last two terms denote the interaction (van der Waals and electrostatics) of the QM and MM region. VdW interactions are always calculated with the MM method, giving the QM atoms vdW parameters. The electrostatic interactions can be included in the MM calculation (mechanical embedding) via definition of atomic charges for the QM atoms and interaction with all MM atomic charges. Another way is to include the electrostatic interactions in the QM calculation (electrostatic embedding). The first method has several disadvantages. In every QM calculation the atomic charges of the QM atoms have to be calculated and the polarisation of the QM region by the external atomic charges is completely neglected in the mechanical embedding scheme

For the QM calculation of the QM region a saturated molecule is needed. If a covalent bond is divided by the QM–MM boundary, the QM fragment lacks an atom to be saturated. The normal way is to put a link atom (L) close to the MM host atom (M_1) and include it into the QM region (Figure 7). Usually this will be a hydrogen atom if an apolar bond is cut.

In the MM calculation the link atom is not included and also nearly all terms with one or more QM atoms are to be deleted. The only terms that are not deleted are the bond term between M_1 and Q_1, angle terms involving M_2, M_1, and Q_1, torsional angle terms involving M_3, M_2, M_1, and Q_1 or M_2, M_1, Q_1, and Q_2 to prevent the QM region to "fly away" or artificially rotate in a geometry optimization.

The link atom is very close to the MM host atom and the original atomic charge on the MM host atom would desturb the QM region artificially large. Therefore,

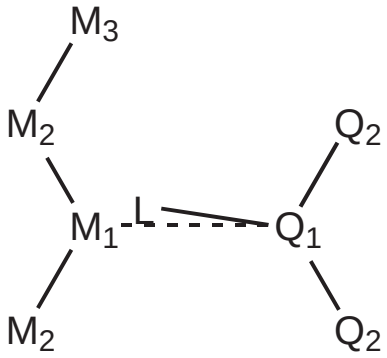


Figure 7: Insertion of a link atom L in the bond between the MM host atom M_1 and the first QM atom Q_1.

the charge on the MM host atom has to be deleted. To not produce a net charge on the host group several link atom schemes were developed. The simplest is to delete all charges on the host group (exgroup, group denotes the group of atoms with neutral charge given in the topology file). In this scheme, maybe important dipole moments are deleted. The second scheme is to shift the deleted charge on all neighboring atoms M_2 equally (shift). This can also change a dipole moment substantially. The third scheme is the *divided frontier charge* linking scheme (DIV). The charge from the MM host atom is divided equally on all atoms of the MM host group. Table 3 gives the both most common cases for a QM–MM boundary and the resulting charges with different link atom schemes.

Table 3: Atomic charges of atoms close to the QM–MM boundary with different link atom schemes.

	orig	exgroup	shift	DIV
boundary between C_α and C_β				
C_α	0.07	0	0	0
H_α	0.09	0	0.113	0.113
N	-0.47	0	-0.447	-0.447
H_N	0.31	0	0.31	0.333
C	0.51	0.51	0.533	0.51
O	-0.51	-0.51	-0.51	-0.51
boundary between C_β and C_γ (e.g. lysine)				
C_α	0.07	0.07	0.01	0.07
C_β	0	0	0	0
$H_{\beta1}$	0.09	0	0.03	0
$H_{\beta2}$	0.09	0	0.03	0

For a QM–MM boundary between C_β and C_γ atoms, the link atom schemes are very similar and the choice does not matter because all involved atoms are nearly uncharged. But for a QM–MM boundary between C_α and C_β the choice is very important. Different schemes were tested by König et al.[27] and the DIV linking scheme is used in this work because both dipole moments between C and O and between N and H_N are conserved.

QM–MM boundary should be in a nonpolar bond to have a realistic description of the bonding situation with the hydrogen link atom and the MM host atom should not be part of a large dipole that would be deleted due to the link atom scheme. In this work all QM–MM boundaries are between C_α and C_β atoms. Thus, only side chain atoms are included in the QM region.

QM/MM calculations for excitation energies are performed in this work in a single-pinot style. The geometry is fixed and thus the bonding and non-bonding terms of the MM region can be neclected for excitation energies. Also the QM/MM VdW interactions are not calculated. Only the interaction of the MM atomic charges with the QM region is taken into the calculation.

2.1.2.1 QM Regions Used in This Work

In this work several different QM regions were used. A very large QM region would be best to give the most exact results. If the chromophore and all surrounding residues would be in the QM region, charge transfer (CT) and polarization effects would be described with the QM method. The MM method has the disadvantages that the charges are fixed and thus no CT and polarization can be in the MM region. But a large QM region is computationally too demanding and thus, a balanced choice of computational speed and accuracy has to be found.

Early works on retinal proteins worked with small QM models of retinal with only a part of the conjugated π system. With the QM methods used in this work, a QM description of the full retinal π system and the β ionone ring is possible and no problems arise from a QM/MM boundary between these two parts. The only QM/MM boundary has to be introduced somewhere between the PSB and the Lys296 backbone C_α atom. The small QM region used has the cutting between the C_α and C_β atoms farthest possible from the PSB. This QM region is called **QM1 region**, is positively charged, and consists of 63 atoms (186 electrons). Smaller QM regions (**QM0a** to **QM0d**) are tested and defined in Section 3.2.

A medium sized QM region is the **QM2 region** consisting of the QM1 region and all sidechain atoms of the counterion Glu113. It is electrically neutral and consists of 73 atoms (226 electrons). It is expected to be much better than QM1 because a charge transfer between the PSB and the counterion is allowed and polarization effects on the Glu113 carboxylate group are important. There are some residues having hydrogen bonds to the charged counterion. Also for these residues although they are neutral, a small charge transfer is expected and also polarization effects are expected to be important because the close interaction with a charged residue. Glu113 is hydrogen bonded by Wat2b, Thr94 and the backbone N–H group of Cys187. To include the backbone atoms is complicated because two new QM/MM boundaries would have to be introduced. This is not tested and the large **QM4 region** consists of QM2 and Wat2b and the sidechain atoms of Thr94 giving a total of 85 atoms (262 electrons). This QM region is the standard region for most of the calculation and if not otherwise noted, results

are for the QM4 region. In Chapter 5 residue 94 is mutated into a leucine in some mutants. No charge transfer is expected to occur between the counterion and the sidechain of Leu94 because of the missing hydrogen bond. But it is still included in the QM4 region to avoid problems in nomenclature by introducing a new QM3 region.

For some calculations, even larger QM regions with an extra residue are used. These regions are called **QM5_x regions** and are composed of the QM4 region and the residue x. In Chapter 5 in many mutants, Gly90 is mutated into a serine that has a hydrogen bond to Wat2b or Glu113. Thus, the **QM5_90 region** is often used there to have a good description of the complex counterion.

2.1.3 Quantum Mechanics

The term quantum mechanics (QM) stands for computational methods that describe molecules in a more exact way. Electrons are explicitly calculated with quantum chemical methods. Different approaches exist and several approximations can be made. This section gives only a very brief introduction. Most of the methods are standard methods (HF, DFT) will be briefly summarized to give the mathematical introduction to explain the approximations made for semi-empirical methods (OM2, SCC-DFTB). Advanced methods called post-HF are described in slightly more detail in Section 2.1.3.3

Common to all these methods is that they base on the time-independent and non-relativistic Schrödinger equation (SE).

$$\hat{H}\Psi = E\Psi \tag{4}$$

Relativistic effects play a non-neglegible role only for heavy atoms (>Kr) and can thus be neglected for biological purposes.[28] This work focuses on stationary points and not on quantum dynamics. Thus, the time-dependent Schrödinger equation has not to be used.

Nuclei can be seen as fixed in space and the electrons moving around them because the velocity of an atomic nucleus is much smaller than of the electrons. This approximation is called the Born-Oppenheimer approximation. Thus, the time-independent SE can be rewritten for the electronic part and the Hamiltonian $\hat{H}$ is written as

$$\hat{H}\Psi^{\mathrm{e}} = E\Psi^{\mathrm{e}} \tag{5}$$

$$\hat{H} = \sum_{i=1}^{N} -\frac{1}{2}\Delta_i + \sum_{i=1}^{N}\sum_{j\neq i}^{N}\frac{1}{r_{ij}} + \sum_{i=1}^{N}\sum_{A}\frac{Z_A}{|r_i - R_A|}. \tag{6}$$

The first term is the kinetic energy of the electrons, the second are the elctron-electron interactions and the last one are the interaction between the nuclei with charge Z_A with all electrons.

To get the total energy, the nucleus–nucleus interactions ($\sum_A \sum_{B \neq A} \frac{Z_A Z_B}{|R_A - R_B|}$) have to be added, that are independent of the electronic wavefunction Ψ^e (called only Ψ from now on).

The task for a high-level QM method is to calculate an approximate wave function that approaches the exact one, and an acurate energy corresponding to this wave function.

2.1.3.1 Hartree-Fock (HF)

In Hartree Fock (HF) theory the N-electron wave function Ψ is written as a single determinant Φ (Slater determinant) consisting of N spin orbitals. A spin orbital χ is a spatial orbital ψ times a spin function α or β. In all methods used in this work, the Hamiltonian is spin-independent, and thus, the spatial orbitals are used. It has to be kept in mind that two electrons can be in every spatial orbital. If the different spins play a role, they are distinguished by a bar (ψ and $\overline{\psi}$).

In contrast to the Hartree theory with a Hartree product

$$\Psi = \psi_1(1)\overline{\psi}_1(2)\psi_2(3)\cdots\overline{\psi}_{\frac{N}{2}}(N), \tag{7}$$

the Slater determinant

$$\Psi = \Phi_0 = \frac{1}{\sqrt{N!}} \begin{vmatrix} \psi_1(1) & \psi_1(2) & \psi_1(3) & \cdots & \psi_1(N) \\ \overline{\psi}_1(1) & \overline{\psi}_1(2) & \overline{\psi}_1(3) & \cdots & \overline{\psi}_1(N) \\ \psi_2(1) & \psi_2(2) & \psi_2(3) & \cdots & \psi_2(N) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \overline{\psi}_{\frac{N}{2}}(1) & \overline{\psi}_{\frac{N}{2}}(2) & \overline{\psi}_{\frac{N}{2}}(3) & \cdots & \overline{\psi}_{\frac{N}{2}}(N) \end{vmatrix} = |\psi_1\overline{\psi}_1 \ldots \psi_{\frac{N}{2}}\overline{\psi}_{\frac{N}{2}}\rangle \tag{8}$$

satisfies anti-symmetry requirements and thus the Pauli exclusion principle by changing sign upon exchange of two electrons. The spatial orbitals (also called molecular orbital, MO) ψ_1 to $\overline{\psi}_{\frac{N}{2}}$ are singly occupied. Other orbitals are unoccupied (virtual MOs). The highest occupied and the lowest unoccupied MOs have special names: HOMO and LUMO.

The MOs are built from a basis set. This is a group of wave function that give the MOs by a linear combination. Usually the basis set consists of atom-centered wave functions that are similar to the atomic orbitals (AOs) of the special atom. The linear combination of atomic orbitals is called LCAO.

$$\psi_i = \sum_j c_{ij}\phi_j \tag{9}$$

The larger a basis set is (the more fuctions are included) the more flexible and more realistic are the MOs and the total wave function. More details on Hartree-Fock theory, the calculation of the HF total energy, the methodology on finding

the best wave function (self-consistent field, SCF) can be found in standard textbooks on theoretical and computational chemistry (e.g. by Szabo and Ostlund,[29] Jensen[28]).

The result of a HF calculation are the basis set coefficients (weighting factors for the linear combination) c_{ij}, energies of all MOs ϵ_i, and the total energy E^{HF}.

2.1.3.2 (HF-Based) Semiempirical Methods

During calculation of the HF total energy, many integrals have to be evaluated, which is the most time-consuming part. It was shown that some of these integrals can be neglected or empirical parameters can be introduced, approximating these integrals, without a substantial loss in accuracy but a considerable improvement in computational speed.

The only HF-based semiempirical method used in this work is the Orthogonalization Method 2 (OM2[1]), which is described in more detail in Section 2.2.3.2.

2.1.3.3 Post-HF Methods

HF theory is not applicable to problems where electron correlation plays a key role. Electron correlation is the effect of the interaction of two electrons avoiding each other because of the same negative charge and the same or different spin. Electrons do not "move" independently but their movement is correlated to the movement of all other electrons in the molecule. Only a part of electron correlation is included in HF theory. The correlation of electrons with different spin is missing and advanced methods based on HF theory (post-HF methods) are needed to account for the inaccuracy of HF theory. The exact energy is always lower than the HF energy. The difference

$$E^{\mathrm{corr}} = E^{\mathrm{exact}} - E^{\mathrm{HF}} \tag{10}$$

is often called correlation energy, although a part of the correlation energy is already included in E^{HF}. The difference is very large for molecules with energetically close-lying occupied and virtual MOs. But also for molecules with large HOMO–LUMO separation, the error in energy and also in an optimized geometry can be large and more sophisticated methods should be used.

HF calculations only give informations about the electronic ground state. To calculate excited states post-HF methods are needed. Configuration interaction is used in this work (see Section 2.2.3). The second class of post-HF methods is perturbation theory. In these methods the Hamilton operator $\hat{H}$ in the SE is written as

$$\hat{H} = \hat{H}_0 + \lambda \hat{H}_1 \tag{11}$$

being the sum of an unperturbed operator $\hat{H}_0$ and a perturbing operator $\hat{H}_1$.

The introduced parameter λ allows the expansion of the energy and the wave function into a series.

$$E(\lambda) = E^{(0)} + \sum_k \lambda^k E^{(k)} \tag{12}$$

$$\Psi(\lambda) = \Psi^{(0)} + \sum_k \lambda^k \Psi^{(k)} \tag{13}$$

Inserting Equations 11 – 13 in the SE, and sorting terms with same exponents of λ gives a set of equations that can be solved iteratively. Usually only the energy in second order is calculated because no perturbed wave function is needed. A simple sum of terms gives the perturbation energy. In Møller-Plesset perturbation theory in second order (MP2), the ground state represented by the HF-Slater determinant Φ_0 is perturbed by the perturbation operator $\hat{H}_1 = \frac{1}{r_{12}}$. The energy in zeroth order is the HF energy, the energy in first order is zero and the energy in second order is given by

$$E^{(2)} = \sum_n \frac{\left\langle \Phi_0 \middle| \frac{1}{r_{12}} \middle| \Phi_n \right\rangle}{E_0^0 - E_n^0} \tag{14}$$

with excited determinants Φ_n contributing to the perturbation energy.

Perturbation theory in second order can also be applied to excited states calculated with a CI method (see Equation 29 in Section 2.2.3.3). In that way the electron correlation is approximated by a perturbation for all electronic states and the energy differences of these states are much more exact than with normal CI methods.

Multiconfigurational self consisting field (MCSCF, CASSCF) calculations can also be corrected by a perturbation calculations in second order. This method is called CASPT2. Excitation energies calculated with this method take a long time and give the best results for medium-sized systems with around 100 atoms, today.

2.1.3.4 DFT

Density functional theory is based on the two important theorems by Hohenberg and Kohn.[30] The first theorem states that the ground state properties of a many-electron system are determined by the electron density $\rho(r)$ that depends on only 3 spatial coordinates instead of $3N$ coordinates for wavefunction based methods (HF, MP2, CI). The second theorem states that the ground state energy is defined by an energy functional and that the correct ground state electron density ρ_0 minimizes the energy calculated with the functional.

The exact functional is not known but it can be in priciple written as a sum of the kinetic energy (functional) of a non-interacting electron gas $T^S[\rho(r)]$, the nucleus–electron interaction $E^{\mathrm{Ne}}[\rho(r)]$, the classical part of the electron–electron interaction $E^{\mathrm{J}}[\rho(r)]$, the exchange–correlation energy $E^{\mathrm{xc}}[\rho(r)]$ (including the correction

to the kinetic energy of the interacting electrons), and the Nucleus–nucleus interactions E^{NN}.

$$E^{\mathrm{DFT}}[\rho(r)] = T^{\mathrm{S}}[\rho(r)] + E^{\mathrm{Ne}}[\rho(r)] + E^{\mathrm{J}}[\rho(r)] + E^{\mathrm{xc}}[\rho(r)] + E^{\mathrm{NN}} \tag{15}$$

with

$$E^{\mathrm{Ne}}[\rho(r)] = \int V^{\mathrm{Ne}}(r)\rho(r)\mathrm{d}r = -\sum_A \int \frac{Z_A\rho(r)}{|R_A - r|}\mathrm{d}r \tag{16}$$

$$E^{\mathrm{J}}[\rho(r)] = \frac{1}{2}\int V^{\mathrm{J}}[\rho(r)]\rho(r)\mathrm{d}r = \frac{1}{2}\int \frac{\rho(r)\rho(r')}{|r - r'|}\mathrm{d}r\mathrm{d}r' \tag{17}$$

$$E^{\mathrm{xc}}[\rho(r)] = \int V^{\mathrm{xc}}[\rho(r)]\rho(r)\mathrm{d}r \tag{18}$$

$$E^{\mathrm{NN}} = \frac{1}{2}\sum_{A\neq B} \frac{Z_A Z_B}{R_A - R_B|} \tag{19}$$

The correct ground state electron density ρ_0, a functional for the kinetic energy of the non-interacting electron gas $T^S[\rho(r)]$ and the exchange correlation functional $E^{\mathrm{xc}}[\rho(r)]$ are not exactly known and calculatable.

For the first two problems, Kohn and Sham introduced one-electron wave functions $\psi_i(r)$ and a non-interacting system of N electrons in these orbitals give the electron density via

$$\rho(r) = \sum_i n_i \left\langle \psi_i(r)|\psi_i(r)\right\rangle, \tag{20}$$

giving the Kohn-Sham-Equations

$$\left[-\frac{1}{2}\nabla_i^2 + V_{\mathrm{eff}}\left[\rho(r)\right]\right]\psi_i(r) = \epsilon_i\psi_i(r) \tag{21}$$

with the effective potential V_{eff} describing the interaction with the nuclei and all other electrons. The total DFT energy than writes as

$$E^{\mathrm{DFT}}[\rho(r)] = \sum_i n_i\epsilon_i - \frac{1}{2}\int V^{\mathrm{J}}[\rho(r)]\rho(r)\mathrm{d}r - \int V^{\mathrm{xc}}[\rho(r)]\rho(r)\mathrm{d}r + E^{\mathrm{xc}}[\rho(r)] + E^{\mathrm{NN}}. \tag{22}$$

In the local density approximation (LDA) the exchange-correlation functional $E^{\mathrm{xc}}[\rho(r)]$ is given as $\int \epsilon^{\mathrm{xc}}[\rho(r)]\rho(r)\mathrm{d}r$. In gradient corrected functionals (generalized gradient approximation, GGA) the exchange-correlation functional is described as

$$E^{\mathrm{xc}}[\rho(r), \nabla\rho(r)] = \int \epsilon^{\mathrm{xc}}[\rho(r), \nabla\rho(r)]\rho(r)\mathrm{d}r. \tag{23}$$

A widely used GGA-functional is PBE, developed by Perdew, Burke, and Ernzerhof.[31]

Hybrid functionals calculate the exchange part of E^{xc} with the exchange part of the Hartree-Fock formalism. In this work the PBE0 hybrid functional[32] is used.

2.2 Advanced Computational Methods

The previous section gives some basic insight into the methods in computational chemistry. This section gives a deeper insight in the methods used in this work. First the methods to calculate the electronic ground state are given (CHARMM in Section 2.2.1.1, DFTB in Section 2.2.2) and then the methods to calculate electronically excited states (OM2 and SORCI in Sections 2.2.3.2 and 2.2.3.3)

2.2.1 System Setup

Protonation states of all titratable amino acids are in the standard state except Asp83 and Glu122 that are assumed to be protonated. Some structures are built with these residues being charged to support this decision (see Section 4.2.1). A disulfide bridge is formed between the two cysteins Cys110 and Cys187. All buried water molecules present in the crystal structure (44 molecules) are included in the calculations.

2.2.1.1 CHARMM

QM/MM optimizations, MD simulations, C=O vibrational calculations, extraction of snapshots from an MD trajectory, calculations of DFTB Mulliken charges, and mutations were done with the CHARMM program package (version 32).[33] The force field used is the CHARMM27 force field.[34]

The cut-off for non-bonded interaction was chosen to be 13 Å. Optimizations were performed with the adopted basis Newton-Raphson method (ABNR) implemented in the CHARMM package. Optimizations were done with low tolerance of 0.0, and a maximum step number of 10000. For MD sumulations the system was heated and equilibrated for 20 ps with an temperature increment of 10 K every 500 steps. Thus, the heating to 300 K takes 15 ps. The production run does not have a standard length. To test the stability of a structure, about 1 ns was enough. For taking snapshots or building histograms even longer simulations are performed. All MD simulations use a standard time step of 1 fs.

Mutations were performed by simply changing the primary sequence and building the new residue with the `IC BUILD` command. For multiple mutants the coordinates of the mutated residues were read from the single mutant structure file.

The calculation of vibrational frequencies is performed with the `VIBRAN` module included in the CHARMM package. The QM region for these calculations is QM1 or QM4 for C=C and C=N stretch vibrations and QM5_x, with the residue x added to the QM4 region, for C=O stretch frequencies of Asp83, Glu122, Glu181, and Gln181.

2.2.1.2 Geometrical Restraints

Structures were set up starting with the crystal structure (pdb: 1U19). Hydrogen atoms were added with the `HBUILD` module included in the CHARMM package. After a short minimization to avoid artificial overlapping of hydrogen atoms, the coordinates are saved and the positions of the non-hydrogen atoms are used as reference position for geometrical restraints of far away atoms. The restraining force is calculated with the B-factors of the crystal structure via

$$f = 0.5 \cdot \frac{3}{2} RT \cdot \frac{8\pi^2}{3B} \tag{24}$$

with $T = 300\,\mathrm{K}$. All atoms closer than $r_1 = 6\,\text{Å}$ to retinal, Thr94, Glu113, and Glu181 are not restrained to give a free mobile region. Atoms with a distance larger than $r_2 = 15\,\text{Å}$ from this region are restrained with the full force constant Atoms in between have a scaled force constant dependant on their distance. The scaling factor is calculated via

$$s(r) = \frac{(r - r_1)^2 \cdot (3 \cdot r_2 - r_1 - 2r)}{(r_2 - r_1)^3}. \tag{25}$$

2.2.1.3 Charge Scaling

All calculations are performed without water and membrane surrounding the protein. The structurally stabilizing effect of the membrane is mimiced with the geometrical restraints of non-hydrogen atoms. The electrostatically shielding effect of the bulk water around the cytoplasmic and extracellular side is modeled with a scaling procedure of charged residues, as proposed by Dinner et al.[35] The movement of the cytoplasmic and extracellular loops is not important, thus, a scaling should produce charges that are realistic in the interaction with retinal and the counterion.

For the calculation of the scaling factors, a set of Poisson-Boltzmann calculations were performed for every charged residue, once in vacuo and once with the dielectric constants of the membrane and protein, set to $\epsilon_1 = 2$, and water, set to $\epsilon_2 = 80$.[36] The protein is oriented perpendicular to the x–y plane and for a inner region (membrane) ϵ_1 and for the outer regions ϵ_2 is used. The scaling factor is calculated to mimic the screening effect of the second PB calculation.

Glu113 is not scaled because it is very close by and is included in the QM4 region. Glu181 is also not scaled because it very close by, too. For all other charged residues the side chain atoms are scaled down. The scaling factors range from 0.00012 (Lys16, 33 Å from retinal) to 0.15 (Asp190, 13 Å from retinal in loop E-II).

The scaled charges are used in all calculations: optimizations, MD simulations, vibrational analysis and excited state calculations.

2.2.2 DFTB

Density functional based tight binding (DFTB[37]) is a semiempirical method based on the DFT total energy equation (Eq. 22). The functionals $V^{\mathrm{J}}[\rho]$, $V^{\mathrm{xc}}[\rho]$, and, $E^{\mathrm{xc}}[\rho]$ are expanded in a Taylor series in second order around a reference density ρ^0. A minimal basis is used. Hamilton matrix elements with respect to the reference density $H^0_{\mu\nu}$ and a repulsive potential E^{rep} are tabulated. In DFTB2, charge fluctuations between atoms are included giving a second order term depending on the atomic charges q and a distance dependant function γ. The DFTB2 [1] total energy is

$$E^{\mathrm{DFTB2}} = \sum_{i\mu\nu} n_i c^*_{\mu i} c_{\nu i} H^0_{\mu\nu} + \frac{1}{2} \sum_{AB} \Delta q_A \Delta q_B \gamma_{AB} + E^{\mathrm{rep}}. \tag{26}$$

2.2.2.1 Improvements of DFTB (DFTB3–diag)

DFTB2 fails for properties of molecules with highly localized charges because in the second order term the charge-charge interactions are independent of the charge state. Indroducing this dependence gives a correction term that includes the third derivative of the atomic energy by the atomic charge $U^{\mathrm{d}}_A = \frac{\partial^3 E^{\mathrm{atom}}}{\partial q_a^3}$. This parameter is called the Hubbard derivative as the second derivative is the Hubbard parameter U_A (describing the chemical hardness of atom A). The DFTB3-diag total energy then is written as

$$E^{\mathrm{DFTB3-diag}} = \sum_{i\mu\nu} n_i c^*_{\mu i} c_{\nu i} H^0_{\mu\nu} + \frac{1}{2} \sum_{AB} \Delta q_A \Delta q_B \gamma_{AB} + \frac{1}{6} U^{\mathrm{d}}_A \Delta q^3_A + E^{\mathrm{rep}}. \tag{27}$$

The Hubbard derivatives have been fitted to reproduce proton affinities and hydrogen-bonding binding energies in the publication of Yang et al.[39] The used Hubbard derivetives are -0.14, -0.09, -0.08, and -0.08 for oxygen, nitrogen, carben, and hydrogen, respectively.

2.2.3 Calculation of Excitation Energies (Configuration Interaction)

To calculate excitation energies, informations of the electronic ground state as calculated with HF, MP2, DFT, or DFTB are not enough. Informations about the electronically excited states, at least the energy, are needed. Chemists often think that calculating an excited state is simply putting one electron from

[1] DFTB2 was formerly called self-consistent-charge DFTB (SCC-DFTB). The name DFTB2 has very recently been introduced by Gaus, Cui, and Elstner.[38] In their work, the complete third order expansion is introduced to DFTB, called DFTB3. The DFTB3–diag method was introduced by Yang et al.[39]

HOMO into LUMO and then again calculating the energy. This is wrong in two points. First, all other orbitals would change when changing one electron and second, the excited state is not simply described by this one slater determinant $|\phi_1^2\phi_2^2\ldots\phi_{\mathrm{HOMO}}^1\phi_{\mathrm{LUMO}}^1\rangle$. Also the description of the ground state is not exact with a single determinant $|\phi_1^2\phi_2^2\ldots\phi_{\mathrm{HOMO}}^2\rangle$. All states Ψ_i are built as linear combinations of many determinants (also called configurations)

$$\Psi_i = \sum_j a_{ij}\Phi_j \tag{28}$$

This ansatz is called configuration interaction.

When solving the Schrödinger equation with this expansion into many determinants, the so called CI matrix with matrix elements of the type $\langle\Phi_\mu|\hat{H}|\Phi_\nu\rangle$ has to be diagonalized. The number of possible configurations rises very fast with increasing number of MOs. A Full-CI calculation is only possible for small molecules and small basis sets. For larger molecules and/or larger basis sets a truncated CI method has to be used. Truncated means not all configurations are included in the CI space. The selection of configurations is a very important point and depends on how accurate the results should be. Details on CI theory can be found in basic textbooks, e.g. Szabo and Ostlund.[29]

For exact calculations of ground state properties, a truncation of the CI space on double excitations (CID) is sufficient. Calculations of excited states need single excitations (CIS). Basic properties of simple molecules (e.g. transition dipole moments) are sufficiently well described with CIS, but for excitation energies higher excited configurations have to be also included (CISD and higher). The main problem for calculating excitation energies is the balanced description of the correlation energy of the states of interest. The complete correlation energy can only be calculated with a full CI calculation and thus a balanced description would be that for both states a similar amount of energy is missing to the exact energy.

Many different methods to truncate the CI space are known. Excitations of electrons from low-lying orbitals or into high-lying orbitals can be neglected. Core electrons of heavy atoms and sometimes even unimportant valence orbitals can be frozen (frozen MOs). With large basis sets, many virtal orbitals are very diffuse and very high in energy. Also these orbitals can be neglected (neglected MOs). The remaining orbitals are called active orbitals. Truncating the degree of excitation on single, double, triple, quadruple, etc. excitations relative to the HF determinant are standard truncating methods. Advanced methods are to allow all excitations in an active space and take all configurations built by these excitations. These methods are called complete active space (CAS). A similar method is the restricted active space (RAS). Therefore, the active space is divided into 3 subspaces (RAS1–3). In the center one, all excitations are allowed and additionally excitations up to a specific level are allowed in RAS1 and RAS3. A short notation is RAS(8:3 2/2/3 2) which means the RAS consists of 8 MOs in total of which 3 are in RAS1, 2 in RAS2 and 3 in RAS3. In RAS2 all excitations are allowed

and additionally up to two excitations from RAS1 (holes) and up to two excitations into RAS3 (particles) are allowed. Examples for allowed configurations are $|2222\,0000\rangle$, $|2220\,2000\rangle$, $|1222\,0010\rangle$, $|2120\,2001\rangle$, $|2112\,2000\rangle$, and $|2020\,2002\rangle$. In general it is a CISD method with HOMO and LUMO orbitals being pronounced and thus some excitations up to quadruple are included.

Another problem with complicated molecules is that the HF wave function can be qualitatively wrong. This means the MOs are wrong and because the CI calculation depends on these MOs the calculated excitation energies are wrong, too. MOs are qualitatively wrong if the ground state is badly described by a single determinant. A method to handle with these molecules is multi-configurational self-consistent field (MCSCF). Here the c_{ij} and a_{ij} coefficients to build the MOs from the AOs and to build the wavefunction from many determinants are determined in parallel. It is a very time-consuming but also very acurate method. Another word for MCSCF is CASSCF, where CAS stands for complete active space. Some MOs are selected (active space) and all configurations build by all excitations (complete) within these orbitals are taken in the CASSCF calculation. Performing a PT calculation after a CASSCF calculation is called CASPT2 which is a very time-consuming method, but it gives very acurate absolute energies and excitation energies.

2.2.3.1 Multi-Reference CI (MRCI)

Molecules with energetically-close occupied and virtual MOs usually have ground and excited states that are not well described by a single determinant. Calculating the correlation energy by a CISD calculation is not enough in these cases because other dominant configurations especially of the excited states are not as well correlated as the HF determinant. To get a balanced correlation more than one reference configuration is needed. Onto these references, all single and double excitation are taken into the CI matrix. The choice of references for such a multi-reference CI (MRCI) calculation can be done manually (Φ_0, Φ_1, Φ_2, and Φ_5) or automatically, e.g. important configurations from a previous CI calculation or all configurations formed by a RAS). MRCI calculations can predict excitation energies very accurately.

In this work two different MRCI methods are used: OM2/MRCI and SORCI, described in the next sections.

2.2.3.2 OM2/MRCI (GUGA-CI)

The Orthogonalization Method 2 (OM2) was developed by Weber et al. in the group of Walter Thiel.[1] It is a semiempirical method based on the NDDO (neglect of diatomic differential overlap) approximation. NDDO has the consequence that the calculated MOs are non-orthogonal to each other. Antibonding

MOs are too low and because the LUMO and most of the virtual MOs are antibonding, the HOMO–LUMO gap is strongly underestimated. With the changed Hamiltonian in OM2 the MOs are nearly othogonal and the energies of antibonding MOs are higher. Also HOMO–LUMO gaps are higher.

The improved description consequently improves the results of heats of formation, rotational barriers, relative energies of isomers, and barriers for some reactions.[40] OM2 can be used in combination with MRCI and gives nice results for excitation energies in a very short time.

Theoretical Background

The CI method included in the MNDO program package[41] and used in this work is GUGA-CI. The graphical unitary group approach (GUGA)[42] is a calculation method that uses symmetry, group theory and spin to fasten the calculation of excited states in a CI calculation. The charges of MM atoms can be included to get a QM/MM excitation energy.

An OM2/MRCI calculation starts with a very fast (few seconds) SCF calculation to get the molecular orbitals. Not all MOs are active in the CI calculation to save some computational time. The active orbitals are selected by simple energetic reasons or by an automatic evaluation of the importance of all orbitals, choosing the most important ones.

The calculations were done in a multi-reference fashion. Some references can be chosen manually or (semi-)automatically. Many parameters can be used to control the calculation. The parameters were benchmarked extensively in former works.[43,44] Several parameters are again tested and the results and the choice is given in the following (see Appendix A for more details).

Practical Use and Parameters

The choice of active orbitals was simply to take the energetically highest occupied and lowest virtual orbitals (Keyword: `MOVO=0`) because in some calculations problems with the automatic choice of the most important orbitals (`MOVO=8`) were seen. The size of this active window was chosen to be 32 MOs (16–16, `ICI1=16 ICI2=16`).

The OM2/MRCI calculation is basically a CISD calculation (`LEVEXC=2`) and higher excitations are included due to the multi-reference fashion (`NCIREF>0`). Five references were taken as the starting point for the CISD calculation (`NCIREF=5`). These were found by many calculation with only one starting reference to be the most important configurations. In short notation the references are: `2200`, `2110`, `2020`, `1201`, and `1111`.

After a first CI calculation it is automatically checked, if all states of interest have a sum of weighting factors of all reference configurations being larger than a

threshold value called CISELR. If not, more references are added and the CISD calculation on these enlarged number of references is repeated (MCIREF=4).

Excitation energies calculated with the OM2/MRCI method are generally too high. The difference is dependent on the QM region and the electrostatic environment (see Table A-4 in Appendix A).

2.2.3.3 SORCI

SORCI is an ab initio method to calculate excitation energies. It is a combination of MRCI and MRPT2 methods. Calculations take a much longer time than OM2/MRCI and therefore, the parameters, thresholds and selection of orbitals are very important to get a good relationship of accuracy and computation time. The method was developed by Frank Nesse[3] and is included in the program package orca developed by Neese's group.[2]

Theoretical Background

The computational procedure is described briefly in this work. The full description with all references to the used corrections and theoretical methods can be found in Ref. 3. Details of used basis set, parameters and choice of orbitals are given afterwards.

A SORCI calculation starts with a Hartree-Fock SCF calculation. Improved virtual orbitals are generated.[45] Some orbitals are selected to give the reference configurations. Either manual specification of references, or automatic generation is used. In this work, the references are generated using a restricted active space (RAS). The reference space is diagonalized and unneccesary configurations with a weight of $\leq T_{\mathrm{Prediag}}$ are neglected and the reference space with the surviving references S is rediagonalized to give the 0th order wave function for each state of interest $\Psi_I^{(0)}$, $\Psi_{II}^{(0)}$, $\Psi_{III}^{(0)}$ and so on.

Configurations are built before the next step. Therefore, the MO space is partitioned in 5 subspaces. First are the frozen orbitals. These are doubly occupied in all configurations. The opposite are the neglected MOs that are empty in all configurations. In between are the internal orbitals that are doubly occupied in all references but can be partly occupied in some configurations. The opposite are the external MOs that are empty in all references but can be partly occupied in some configurations. In the very middle are the active MOs that build the RAS and are partly occupied in the references.

The selection of configurations is done with the difference-dedicated configuration interaction (DDCI[46]) scheme. Ref. 3 gives a table of the 14 types of configurations. In the first iteration a small DDCI2 space is used. In later iterations the larger DDCI3 space is used. Both have in common that configurations that are built by an excitation of two electrons from internal into external MOs

($\text{int}^2 \rightarrow \text{ext}^2$) are not included. All remaining single and double excitations in internal, active, and external MOs are included in the DDCI3 space. The difference to DDCI2 are the configurations that are built from an excitation of two internal electrons into an active and an external MO ($\text{int}^2 \rightarrow \text{act}^1 + \text{ext}^1$) or built from an excitation of one internal and one active electron into two external orbitals ($\text{int}^1 + \text{act}^1 \rightarrow \text{ext}^2$) are not included in DDCI2.

For all configurations Φ_μ, the perturbation $E_{I,\mu}^{(\text{pert})}$ is calculated

$$E_{I,\mu}^{(\text{pert})} = \frac{\left|\left\langle \Psi_I^{(0)} \middle| \hat{H} \middle| \Phi_\mu \right\rangle\right|^2}{\left\langle \Psi_I^{(0)} \middle| \hat{H}_0 \middle| \Psi_I^{(0)} \right\rangle - \left\langle \Phi_\mu \middle| \hat{H}_0 \middle| \Phi_\mu \right\rangle}. \tag{29}$$

Configurations with a perturbation larger than a threshold T_{Sel} are selected to be part of the strongly-perturbing subspace R', the others are part of the weakly-perturbing subspace R''. The space consisting of reference configurations S and strongly-perturbing configurations R' is then diagonalized to give new wave functions Ψ_I^{diag} and energies E_I^{diag}. The energies of the states I, II, ... are given as

$$E_I = E_I^{\text{diag}} + \sum_{\mu \in R''} E_{I,\mu}^{(\text{pert})} + E_I^{\text{corr}} \tag{30}$$

with a correction of the Davidson type.

To be independent on the quality and choice of the orbitals in the very first step of a SORCI calculation, the calculation is iterated. Therefore, approximate average natural orbitals (AANOs) are built and used as input for the next iteration. AANOs with an occupation number N less than the threshold T_{nat} or larger than $2.0 - T_{\text{nat}}$ are used as frozen and neglected orbitals. In the second and subsequent iterations the larger DDCI3 space is used to build configurations.

Practical Use and Parameters

The basis set used for all SORCI calculations is SV(P).[47] In calculations with the counterion included in the QM region, a diffuse s and p function is added to the two oxygen atoms of Glu113 (with exponent 0.0845).

Retinal consists of six double bonds and thus twelve π MOs. These are used to build the RAS. A RAS(12:5 2/2/5 2) is used to give them a more pronounced attention than other orbitals. The choice of orbitals for the generation of the reference space S is very crucial. For most of the calculations in this work, a manual selection (rotation) of the SCF MOs was performed. This and other methods are described and benchmarked in Appendix B.

For the partitioning of the MOs in the 5 subspaces, energetical thresholds are used. MOs with an energy lower than $-3\,\text{E}_\text{h}$ or higher than $5\,\text{E}_\text{h}$ are kept frozen or are neglected (`ewin -3,5`). The prediagonalization threshold T_{Prediag} is set

to 10^{-3} (`tpre 1e-3`). The selection threshold T_{Sel} is kept very low with $10^{-6}\,\mathrm{E_h}$ (`tsel 1e-6`). The selection threshold for natural orbitals in the second itaration is 10^{-6} (`tnat 1e-6`). Only one iteration is done. The choice of basis set, thresholds and iteration number is from an extensive benchmark work on protonated Schiff base models by Wanko et al.[44] and in the original SORCI paper by Neese.[3]

SORCI excitation energies are converged within 0.05 eV of the method's limit.[44,48] For calculating excitation energies of the lonely retinal PSB without counterion and other external charges there are some problems with the level ordering of higher excited states. This problem is described in detail in Appendix C.

2.2.3.4 Polarization Model

CHARMM charges are used for all CHARMM calculations (optimizations, MD simulations, vibrational analysis). These charges are also used as external point charges for the calculations of excitation energies with SORCI. CHARMM charges are prepolarised to mimic the polarization effect of other amino acids being hydrogen bonded. For some excited state calculations, a polarization model is used. This model has been developed by Marius Wanko and is described in detail in Refs 49 and 50.

As input for the polarization calculations ESP-fitted charges are used. A dipole is added to each of these charges. These dipoles can freely orient and change theirs moment according to their polarizability (polar.h model). The orientation and moments of these dipoles are optimized with the charge distribution of the QM region in the S_0 and S_1 state and by all other charges and dipoles. The ground and excited state (S_0 and S_1) are calculated with the polarized external charges giving two calculations. The external charges are again polarized by the newly calculated stated. This procedure is iterated till the excitation energy is converged. The excitation energy of the calculation with external charges polarized by the S_0 state is called the vertical excitation energy ΔE^{vert}. For the calculation of the polarized excitation energy ΔE^{pol}, also the excitation energy of the calculation with charges polarized by the S_1 state are needed and also the polarization energies of the external charges and dipoles. Approach ii is used in this work.[49]

A large difference of the unpolarized $\Delta E^{\mathrm{CHARMM}}$ and the vertical excitation energy ΔE^{vert} shows the error of the constant CHARMM charges. This can be seen with small QM region, where the polarization model includes the polarization effect of the counterion by the PSB charge. The difference of ΔE^{vert} to the polarized excitation energy ΔE^{pol} is called polarisation energy and includes the difference in polarization due to the changed charge distribution in the QM region in S_0 and S_1 states.

Chapter 3

Color Tuning in Rhodopsin

The term color tuning is used in two different ways in literature and in this work. Chapter 5 studies the color tuning of different retinal proteins (Rh and HB). Electrostatic and structural changes between the two structures determine the difference in excitation energy. This chapter studies the absolute excitation energy of the retinal chromophore in Rhodopsin. This is tuned by the electrostatic interaction of the protein residues with the chromophore (Section 3.1). For calculating absolute excitation energies correct structures are needed and also the method to calculate the excitation energy has to be a high-level one. The method used in this work are DFTB-CHARMM for structure optimizations (see Section 2.2.2) and SORCI for calculating excitation energies (see Section 2.2.3.3). These are sufficiently exact for a largely conjugated π system as retinal.The right choice of QM region used for the calculations is a topics in this chapter, too (Sections 3.2). Charge transfer and the inclusion of polarization effects are important for the decision.

3.1 Electrostatic Effects of Color Tuning

3.1.1 Methods

Charges of atoms in the QM region and dipole moments of the ground and excited states are calculated with the SORCI method. Different models for the electrostatic embedding (EE) are used. For all QM regions, different schemes were used described in Table 4. For the medium-sized QM2 region an extra calculation QM2–QM4 is performed with the point charges of Wat2b and Thr94 (being part of the QM4 region). For the QM1 regian an extra calculation including the point charges of Glu113 (included in QM2) is performed (QM1–QM2).

Table 4: Nomenclature for electrostatic embedding.

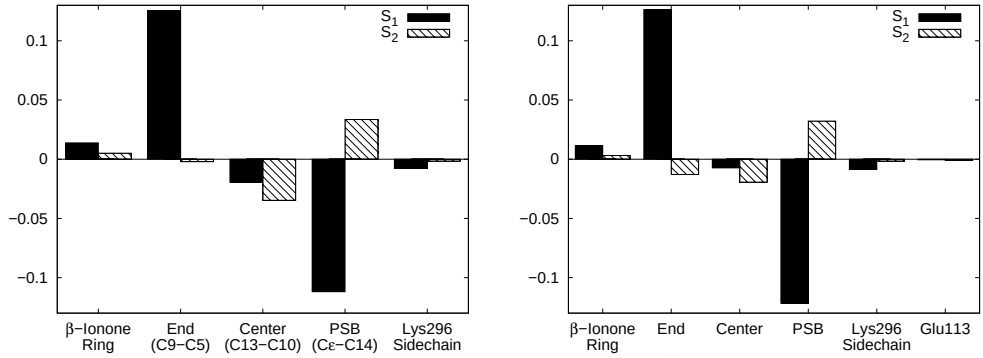
Name	Description
no	Without external charges
QMx	Only with point charges of atoms building the QMx region
CCI	+ point charges of Glu113's C=O group and Cys187's NH group (both having hydrogen bonds to Glu113's carboxyl group and thus building the complex counterion)
HBN	+ point charges of Wat2a, Cys187's C=O group, and sidechains of Ser186, Glu181, Tyr268, and Tyr191 (building the HBN on the extracellular side of retinal)
pert	+ point charges of sidechains of Tyr43, Thr118, His211, Trp265, and Phe293 (these show a large effect on the excitation energy in the perturbation analysis (see Section 5.3.1.2))
28AA	+ point charges of the 28 amino acids building the binding pocket as described by Sekharan et al.:[51] 43, 44, 47, 94, 113, 117, 118, 120, 121, 122, 178, 181, 186, 187, 188, 189, 191, 192, 207, 211, 212, 261, 265, 268, 269, 293, 295, 296, and Wat2a
xÅ	+ point charges of amino acids and water molecules with one or more atoms within x Å around the QM4 region ($x =$ 3: +13 AA, $x = 4$: +16 AA, $x = 5$: +30 AA, $x = 6$: +45 AA +2 Wat, $x = 7$: +64 AA +6 Wat, $x = 8$: +76 AA +8 Wat, $x = 9$: +99 AA +9 Wat.)
all	With point charges of all amino acids and waters

3.1.2 Results

Electrostatic effects play an important role in color tuning of retinal proteins because the excitation from the electronic ground state (S_0) to the first excited state (S_1) contains a large shift of the positive charge from the PSB to the end of the polyene and the β-ionone-ring. Figure 8 shows the amount of charge translocation due to this excitation. This effect is much smaller for the $S_2 \leftarrow S_0$ excitation, where a small charge movement is observable in the other direction.

For the QM2 and QM4 regions a change in dipole moment can be calculated. For the QM1 region a calculation of a dipole moment is not possible because of the net positive charge. But differences in dipole moments can be calculated if the same origin is taken. The state dipole moments $\vec{\mu}$ are very much depending on the QM region and the electrostatic environment (Table 5), but the difference dipole moments $\Delta\vec{\mu}$ (Table 6) are very similar for QM2 and QM4.

From the magnitude $|\Delta\vec{\mu}|$ and the direction of the difference dipole moment $\Delta\vec{\mu}$ the effect of changed electrostatics in the environment of retinal can be estimated. But also the electrostatic environment has an effect on the difference dipole mo-



(a) QM1 region **(b)** QM2 region

Figure 8: Charge translocation (e) on retinal due to $S_1 \leftarrow S_0$ and $S_2 \leftarrow S_0$ excitation (Calculations with all point charges).

Table 5: State dipole moments $\vec{\mu}$ (D) of the S_0 state with different QM regions and electrostatic embedding.

	EE	μ_x	μ_y	μ_z	$\lvert\vec{\mu}\rvert$
QM1[a]	QM4	*-0.36*	*20.11*	*9.45*	*22.22*
	all	*-1.18*	*18.87*	*10.54*	*21.65*
QM2	QM4	6.64	-10.47	-6.63	14.06
	all	5.91	-11.64	-5.79	14.28
QM4	no	8.11	-12.61	-2.86	15.26
	all	7.66	-14.53	-2.08	16.56

[a] Dipole moments for the QM1 region are not trustworthy because they depend on the choice of the origin.

Table 6: Difference dipole moments $\Delta\vec{\mu}$ (D) for $S_1 \leftarrow S_0$ and $S_2 \leftarrow S_0$ excitations with different QM regions and electrostatic embedding.

		EE	$\Delta\mu_x$	$\Delta\mu_y$	$\Delta\mu_z$	$\lvert\Delta\vec{\mu}\rvert$
$S_1 \leftarrow S_0$	QM1	QM4	–1.59	–6.74	–2.74	7.45
	QM2	QM4	–1.62	–6.78	–2.66	7.46
	QM4	no	–1.68	–7.06	–2.75	7.76
	QM1	all	–1.68	–6.44	–2.46	7.10
	QM2	all	–1.78	–6.75	–2.46	7.40
	QM4	all	–1.86	–7.01	–2.54	7.69
$S_2 \leftarrow S_0$	QM1	QM4	–0.72	–1.26	–0.81	1.66
	QM2	QM4	–0.58	–0.64	–0.49	0.99
	QM4	no	–0.57	–0.66	–0.50	1.01
	QM1	all	–0.94	–1.49	–0.74	1.91
	QM2	all	–0.82	–0.96	–0.50	1.36
	QM4	all	–0.86	–1.05	–0.49	1.45

ment. The change between the QM1–QM4 and the QM1–all calculations is very large for the $S_1 \leftarrow S_0$ excitation. This shows a large electrostatic impact of the external charges on the positively charged retinal chromophore with fixed point charges of the complex counterion. For the $S_2 \leftarrow S_0$ excitation for all QM regions teh change in difference dipole moment with and without external charges is seen. But it is also observable that QM2 and QM4 regions give similar results strongly deviation from teh QM1 results.

For the $S_1 \leftarrow S_0$ excitation the difference dipole moment is much larger than for the $S_2 \leftarrow S_0$ excitation, ≈ 7.5 and 1.5 D, respectively. Thus, the dependence on changes in the electrostatics should be much stronger for the $S_1 \leftarrow S_0$ than for $S_2 \leftarrow S_0$ excitation. This is shown in the following.

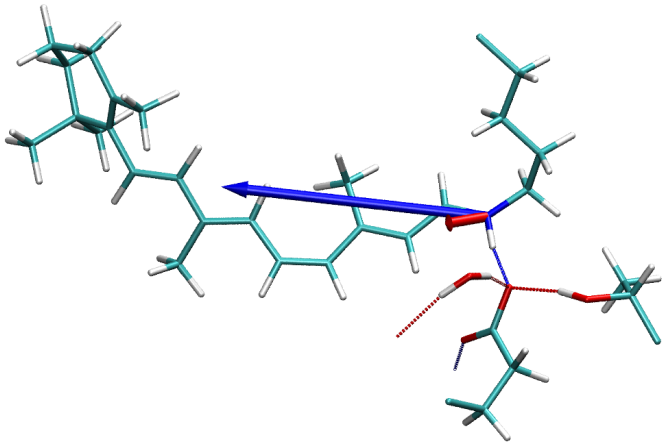


Figure 9: Difference dipole moments $\Delta\vec{\mu}$ for $S_1 \leftarrow S_0$ (blue arrow) and $S_2 \leftarrow S_0$ excitations (red arrow) (QM4-all).

To get the pure effect of the electrostatic environment, the following calculation were all done on the same structure. Successively the point charges of the protein environment are added to the QM1-optimized structure. Additionally the same is done for the QM2 and QM4 region but also with the QM1-optimized structure, where only the new link atoms were added and optimized. The resulting excitation energies for the $S_1 \leftarrow S_0$ and $S_2 \leftarrow S_0$ excitation are shown in Figure 11.

The exact calculation of the S_2 state is computationaly more complicated and thus the difference between the $S_1 \leftarrow S_0$ and $S_2 \leftarrow S_0$ excitation energies are not very exact. In some cases the S_1 state is energetically higher than the S_2. In these cases the assignment of the states to S_1 and S_2 is done by the main configuration Φ_i in the SORCI calculation. It is the $HOMO^1$–$LUMO^1$ configuration for the S_1 and the $HOMO^0$–$LUMO^2$ configuration for the S_2 state. At first view it seems to be wrong that the HOMO→LUMO double excitation should be lower in energy than the single excitation, but e.g. for carotenoids (long polyenes with 9 – 11 conjugated double bonds) it is very common that the state with the HOMO–LUMO single excitation as dominant configuration is the S_2 state. Spectra show that in the case of retinal the state dominated by the $HOMO^1$–$LUMO^1$ configuration is the lowest excited state S_1.

Table 7: SORCI excitation energies for the first iteration and the average of iterations 11 to 20 (QM1). The average excitation energies are scaled down by the average difference.

(a) $S_1 \leftarrow S_0$ excitation energies (eV)

	first iter	11–20 iter	Δ	scaled
no	1.88	2.28	0.40	1.84
QM2	2.68	3.08	0.41	2.65
QM4	2.59	3.03	0.45	2.60
CCI	2.60	3.03	0.42	2.59
hbn	2.68	3.12	0.44	2.68
pert	2.63	3.07	0.44	2.63
28AA	2.49	2.94	0.45	2.50
3Å	2.56	3.00	0.44	2.57
4Å	2.52	2.96	0.44	2.52
5Å	2.54	2.97	0.43	2.54
6Å	2.55	2.99	0.44	2.55
7Å	2.49	2.94	0.45	2.50
8Å	2.50	2.94	0.45	2.51
9Å	2.54	2.98	0.44	2.54
12Å	2.53	2.97	0.44	2.54
15Å	2.52	2.96	0.44	2.52
20Å	2.53	2.96	0.43	2.53
all	2.53	2.97	0.44	2.54
		mean value:	0.44	

(b) $S_2 \leftarrow S_0$ excitation energies (eV)

	first iter	11–20 iter	Δ	scaled
no	2.46	3.13	0.67	2.40
QM2	2.68	3.36	0.69	2.63
QM4	2.65	3.39	0.74	2.65
CCI	2.67	3.36	0.69	2.62
hbn	2.67	3.42	0.75	2.69
pert	2.67	3.42	0.75	2.69
28AA	2.67	3.42	0.75	2.69
3Å	2.70	3.45	0.75	2.71
4Å	2.66	3.44	0.77	2.70
5Å	2.70	3.42	0.73	2.69
6Å	2.70	3.45	0.74	2.71
7Å	2.68	3.43	0.75	2.70
8Å	2.68	3.42	0.74	2.69
9Å	2.70	3.45	0.75	2.71
12Å	2.70	3.45	0.74	2.71
15Å	2.71	3.43	0.73	2.70
20Å	2.70	3.45	0.75	2.71
all	2.70	3.45	0.75	2.71
		mean value:	0.73	

[a] Calculations with no charges were performed with 4 roots in the SORCI calculation because the S_2 and S_3 are different for all other calculations. Excitation energies are ≈ 0.02 eV lower than for 3-root calculations. This shift is the same for all other calculations.

In cases where the S_1 and S_2 states are similar in energy, the resulting excitation energies for the excitation energies are not reliable (see Appendix C). To get a good description for all states of electrostatic embedding, SORCI calculations for the QM1 region were done with 20 iterations. The excitation energies increase with increasing iteration number. After 15 iterations the excitation energies are nearly converged. To average out the fluctuations in the absolute excitation energies and in the shifts, the mean value of the last 10 iterations is taken. This value is scaled down by the mean difference.

The by far largest effect is the shift resulting from the negative charge of the counterion Glu113. Adding the counterion's point charges to retinal in vacuum (no to QM2) gives a blue shift of $+0.81$ eV. The shift of including the QM4 and CCI charges is small (-0.05 eV and -0.01 eV, respectively). Including the HBN

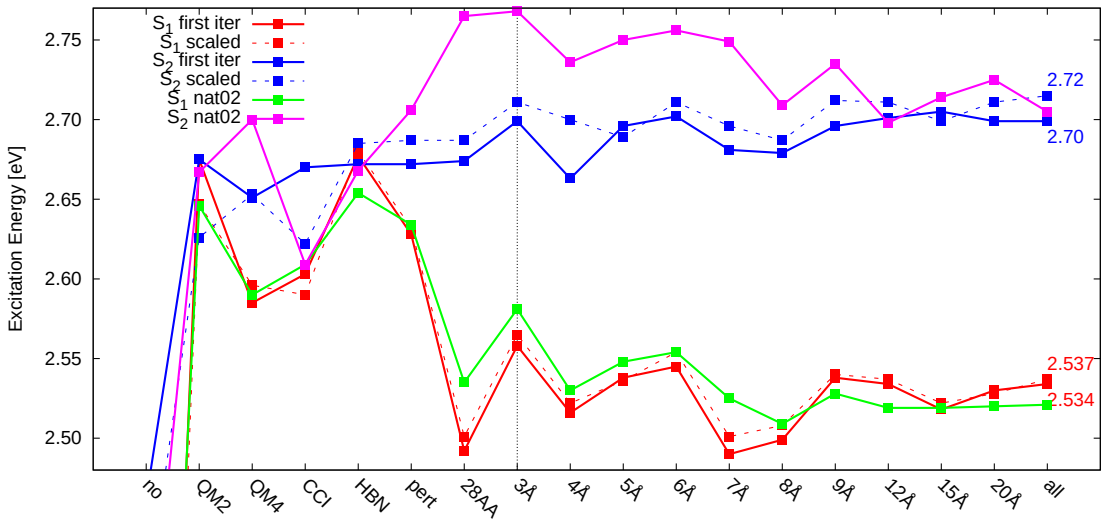


Figure 10: Effect of the electrostatic environment on the $S_1 \leftarrow S_0$ and $S_2 \leftarrow S_0$ excitation energies with QM1 region. SORCI results of the first iteration and scaled values (see Table 7).

charges again gives a bigger blue shift of $+0.09\,\text{eV}$. This is mainly due to the charged Glu181 residue. The added charges of the "pert" residues gives again a red shift of $-0.05\,\text{eV}$. The large red shift ($-0.13\,\text{eV}$)due to the inclusion of the "28AA" residues can be devided into the effect of the missing sidechains or the backbone atoms. Adding only the missing sidechains of the 28 amino acids gives a scaled excitation energy of $2.64\,\text{eV}$ ($+0.01\,\text{eV}$). Adding only the charges of the backbone atoms of all 28 amino acids gives $2.49\,\text{eV}$ ($-0.14\,\text{eV}$). The shifts are nearly additive. Including more and more atoms gives only small shifts.

The shifts for the $S_2 \leftarrow S_0$ excitation are much smaller and difficult to interpret. With only a few external charges (no to cci) the difference between the scaled and unscaled excitation energies are very large. Probably the results of the first few iterations are wrong becaus of any reason. Interpreting the scaled values, there is a much smaller blue shift due to the inclusion of the counterion's charges ($+0.23\,\text{eV}$). The blue shift due to the inclusion of the HBN charges (charged Glu181) is similar to the shift of the $S_1 \leftarrow S_0$ excitation energy ($+0.08\,\text{eV}$). All other shifts are very small. Especcially the strong red shift (pert→28AA) is missing.

For bigger qm regions the SORCI calculation with 20 iterations is too time consuming and it is shown for the QM1 region that discrepancies in the scaled and unscaled values only arise for the S_2 state and for calculations with very few charges. Because of that, the following paragraphs concentrate on the unscaled $S_1 \leftarrow S_0$ excitation energies.

Including the atoms of Glu113 to the QM region (QM1–QM2 to QM2–no) gives a blue shift of $+0.08\,\text{eV}$. This shift is due to the charge transfer $0.072\,\text{e}$ from Glu113 to retinal in the QM2–no calculation. It stabilizes the S_0 state more than the S_1

Table 8: Excitation energies $\Delta E_{S_1 \leftarrow S_0}$ (eV) of the first SORCI iteration.

	QM1	QM2	QM4
no	1.88		
QM2	2.68	2.76	
QM4	2.59	2.69	2.68
CCI	2.60	2.68	2.70
HBN	2.68	2.70	2.70
pert	2.63	2.66	2.67
28AA	2.49	2.56	2.56
3Å	2.56	2.60	2.63
4Å	2.52	2.57	2.57
5Å	2.54	2.59	2.59
6Å	2.55	2.62	2.61
7Å	2.49	2.57	2.55
8Å	2.50	2.58	2.59
9Å	2.54	2.60	2.60
12Å	2.53	2.58	2.56
15Å	2.52	2.59	2.56
20Å	2.53	2.58	2.57
all	2.53	2.59	2.57

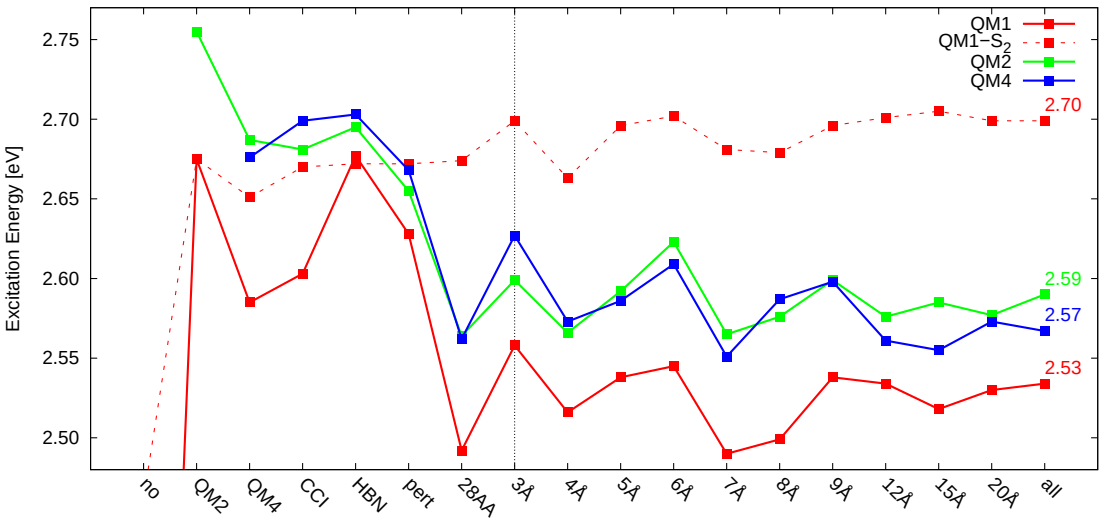


Figure 11: Effect of the electrostatic environment on the $S_1 \leftarrow S_0$ excitation energies with QM1 (red), QM2 (green), and QM4 region (blue). For the description of the different levels of electrostatic embedding see Table 4.

state, where the positive charge on retinal is translocated towards the β-ionone ring. This charge transfer has some impact on the electrostatic dependence of the excitation energy. The red shift QM2–no→QM2–QM4 (-0.07 eV) is similar to the corresponding shift of the QM1 calculations (-0.09 eV but the strong blue shift ($+0.07$ eV in QM1 (CCI→HBN)) due to inclusion of the HBN charges is only $+0.01$ eV for QM2.

Including the atoms of Thr94 and Wat2b to the QM region (QM2-QM4 to QM4-no) gives nearly no shift(-0.01 eV). The shifts when increasing the number of charges are very similar to the QM2 shifts.

Comparing the shape of the 3 curves in Figure 11 in the right part shows very similar behaviour. It is difficult to interpret why e.g. the shift 6Å→7Å is such a stong red shift ($+0.05$ – $+0.06$ eV). The important point is that the shifts add up to nearly zero. The small differences in the shifts come from a different impact of the electrostatics of the different QM regions. To get a correct description of the electrostatic influence of special amino acids (especially close ones) it is very important to choose a big QM region that has a correct description of charge transfer effects and the charge translocation due to electronic excitation.

In summary the color tuning in retinal is mainly dominated by the negative charge of the counterion and some aminoacid backbone atoms. But these effects are not modifiable by mutation of some amino acids. Color Tuning in the sense of fine-tuning the color between different opsin molecules or by single mutations is described by all the remaining small effects. Figure 11 suggests that the overall effect of these effects can only be in the range of ± 0.3 eV.

But the response of the retinal conformation on changes in electrostatics by changing the bond length alternation and twisting of the polyene are not shown in this chapter, because all the calculations were done on similar geometries (QM1, QM2, or QM4 optimized). This indirect effect of electrostatics can only be shown by performing mutations and relaxing the structure.

3.2 Effect of the Choice of the QM Region

As shown in the last section, electrostatics are very important for tuning the excitation energy of retinal. But also the choice of the QM region is important. The size of the QM region is important for the results of the excited state calculation. Shortening the QM region at the lysine side chain has a clear effect on the calculated excitation energy and the save in calculation time is very small (see Table 9, also for nomenclature of QM regions). The large blue shift from QM0a to QM0b ($+0.21$ eV) is because of the enlarged ability of the positive charge to delocalize on the C_ε atom. This stablilizes the S_0 more than the S_1 state. Thus the choice of including the whole sidechain (C_ε, C_δ, C_γ, and C_β) is the best choice.

Enlarging the QM region to the counterion (QM2) or the complex counterion (QM4) includes the polarization of the added atoms and a possible charge trans-

Table 9: $S_1 \leftarrow S_0$ exitation energies with small QM regions (optimized with the same QM region, SORCI, eV), Number of atoms N, calculation time t (h).

		$\Delta E_{S_1 \leftarrow S_0}$	N	t
RET to N_{16}	QM0a	2.27	51	10
RET to C_ε	QM0b	2.48	54	12
RET to C_δ	QM0c	2.54	57	11
RET to C_γ	QM0d	2.51	60	12
RET to C_β	QM1	2.53	63	13

fer to Glu113. These parameters have an effect on the results twice. First in the geometry optimization done with this QM regions and second in the calculation of the excitation energies on these geometries. These effects can not be clearly seperated by each other.

3.2.1 Effects of Structural Changes and Charge Transfer on the Choice of the QM Region

The first effect of the choice of the QM region is a slightly changed structure due to geometry optimization with this QM region. This is quantified in Table 10 by the BLA and the atom distances in the complex counterion. The BLA has a direct effect on the excitation energy and the distances in the complex counterion have an indirect effect due to a change in charge transfer (Table 12).

Table 10: Geometrical parameters (Å) of structures optimized with QM1, QM2, and QM4 region.

	QM1opt	QM2opt	QM4opt
BLA	**0.065**	0.067	0.067
$d(N_{16}–O_{\varepsilon 2})$	**2.77**	2.73	2.70
$d(O_{\varepsilon 2}–O_{wat2b})$	2.66	**2.64**	2.78
$d(O_{\varepsilon 2}–O_{\gamma Thr94})$	2.64	**2.59**	2.69

The effect of the structural differences are:

a) Due to the increased BLA, the excitation energies calculated with the QM1 region without external charges (bare retinal, values in brackets in the first row of Table 11) are blue shifted by $+0.02$ eV.

b) Due to the shortened N_{16}–$O_{\varepsilon 2}$ distance, the charge transfer from the retinal to the counterion (Table 12) is increased. This seems to have a small effect on the excitation energy, because the blue shift for the excitation en-

ergies calculated with the QM2 region without external charges (values in brackets in the second row of Table 11) are only blue shifted by $+0.02\,\text{eV}$.

c) Due to the very short $O_{\varepsilon 2}$–O_{wat2b} and $O_{\varepsilon 2}$–$O_{\gamma\text{Thr94}}$ in the QM2 optimized structure, the charge transfer from wat2b and Thr94 to Glu113 (CT2 in Table 12) is drastically increased. This is the reason for the very high excitation energy of the QM2 optimized structure calculated with QM4 region without external charges ($+0.05$ and $+0.03\,\text{eV}$ relative to the QM1 and QM4 optimized structures.

Table 11: $S_1 \leftarrow S_0$ exitation energies (QM1, QM2, and QM4 region, SORCI, eV) optimized with QM1, QM2, and QM4 region. Calculated with (without) all external charges.

	QM1opt	QM2opt	QM4opt
QM1	2.53 (1.88)	2.48 (1.90)	2.46 (1.90)
QM2	2.59 (2.75)	2.56 (2.77)	2.56 (2.77)
QM4	2.57 (2.68)	2.55 (2.73)	2.50 (2.70)

Table 12: Charge transfer of structures optimized with QM1, QM2, and QM4 region (e).

	QM1opt	QM2opt	QM4opt
CT(QM2)[a]	0.054	0.067	0.071
CT(QM4)[a]	0.051	0.063	0.068
CT2(QM4)[b]	–0.061	–0.073	–0.058

[a] CT is the charge transfer from retinal to the counterion or to the complex counterion $(1 - q(\text{retinal}))$.

[b] CT2 is the charge transfer from wat2b and Thr94 to Glu113 $(q(\text{wat2b}) + q(\text{Thr94}))$.

3.2.2 Effect of Protein Polarization on the Choice of the QM Region

Besides the charge transfer the second effect of increasing the QM region is the added polarisation of the added atoms. As mentioned in Section 2.2.3.4, the charges included in the CHARMM force field are prepolarized to give the mean effect of the atoms being in a normal environment. Especially close to the net charges of the PSB and the counterion, the prepolarized charges are expected to be wrong. Therefore, the effect of explizit polarisation is investigated in the following.

As can be seen in Figure 12 the blue shift due to adding the QM2 charges to the QM1 region is by far the largest ($\approx +0.8\,\text{eV}$). Including the explizit polarization of the QM2 charges (without all other external charges) gives a slight blue

shift of +0.01 eV. The larger effect of increasing the QM region to QM2 comes from the charge transfer of 0.054 e from retinal to Glu113 (+0.04 eV). The relaxation of the structure due to optimization with the QM2 region only gives the small shift of +0.02 eV. Adding the QM4 charges in the QM2 calculation gives a large red shift of −0.08 eV because of the electrostatic effect on retinal (see Figure 11: −0.08 eV, QM1–QM2→QM1–QM4) and the decreased charge transfer (from 0.087 to 0.076 e). Inclusion of the explicit polarisation of the QM4 charges has no effect on the excitation energy. This shows that the prepolarized CHARMM charges are good for water and Thr94. The slight blue shift (+0.04 eV) due to the QM description of the whole QM4 region is due to the allowed charge transfer from Wat2b and Thr94 to the counterion. The effect of the structural relaxation is small (−0.02 eV). A large red shift comes from the external charges around the QM4 region (−0.12 eV), as was shown in Section 3.1. The polarisation effect of these external charges is a large bathochromic shift, too (−0.07 eV).

This shows that polarization effects are very important for calculating absolute excitation energies. For shifts in excitation energy, the polarization effect is not important, because it is similar for different structures and thus cancels out.

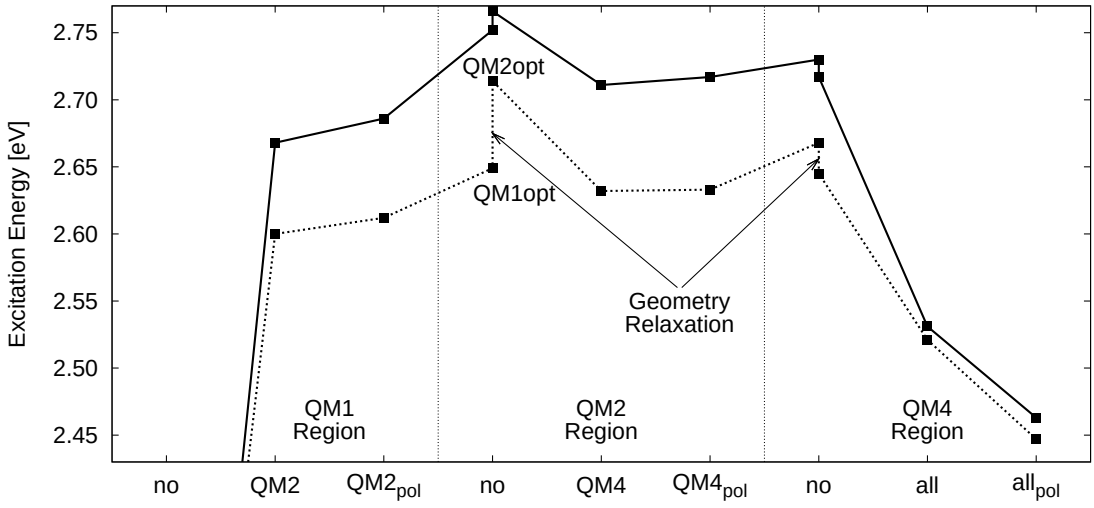


Figure 12: Effect of adding the complex counterion charges, polarization, increasing the QM region, structural relaxation and all other external charges and their polarization.

Figure 13 shows the polarization effect of the complex counterion residues and all external charges on different QM regions. The effect of having the QM4 residues with the explicit polarization model is very different in QM1 and QM2 (−0.02 and +0.02 eV, respectively). Also the polarization effect of the other external charges is different for QM1, QM2, and QM4 region (−0.02 and +0.02 eV, respectively). These differences show that the choice of the QM region is very important. The big QM4 region has strong charge transfer effects that have a structural effect and also an effect on the polarisation of the surrounding residues. To

further increase the QM region could be important but is hard to do because of the computational time.

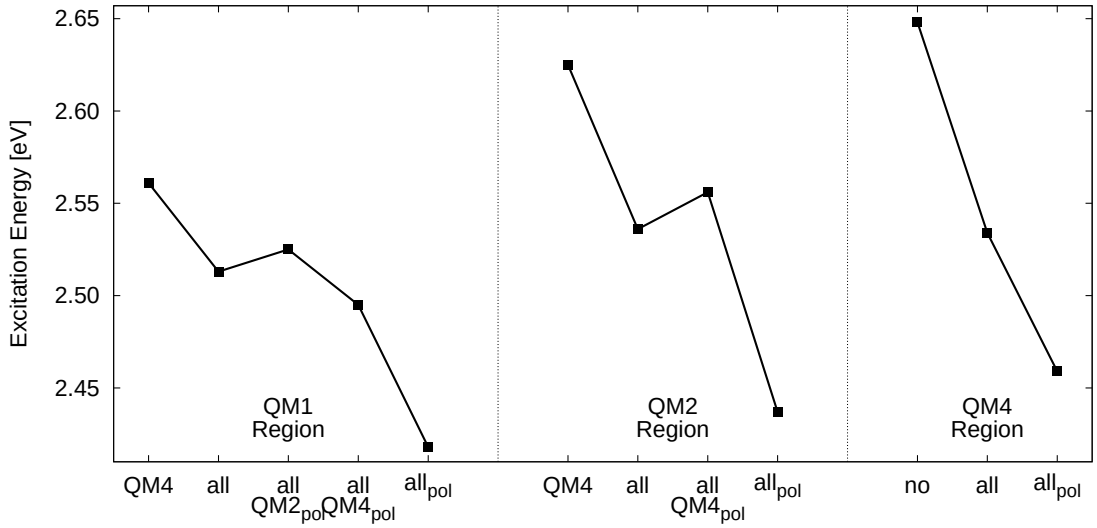


Figure 13: Effect of adding all external charges to QM1–QM4, QM2–QM4, and QM4–no systems. Polarization effect of the complex counterion residues and all charges.

3.2.3 Summary

The QM4 region is chosen to be the standard QM region because the charge transfer from the PSB to the counterion is converged and the important polarization effects on the complex counterion are included in the QM region. The polarization effect of the remaining residues is relatively small and is thought to be similar for the mutants in Chapter 5 and thus, it is sufficient for calculating shifts in excitation energies. For absolute excitation energies, the polarization model is thought to enhance the accuracy of the calculations with the QM4 region sufficiently and no enlarged QM regions are needed except for some mutants including the E90S mutation, that is close to the complex counterion.

Chapter 4

The Problem of Glutamic Acid 181 (Glu181) in Rhodopsin

4.1 Introduction to the Problem of Glu181 Protonation State

4.1.1 Crystal Structures

Glutamic acid 181 (Glu181) was mentioned with respect to its protonation state in Section 1.4.3. It is very close below retinal (6 Å) and is part of a HBN located parallel to retinal. The crystal structures are not sure about the structure of the HBN and of course the protonation state of Glu181 can not be determined with this method, too.

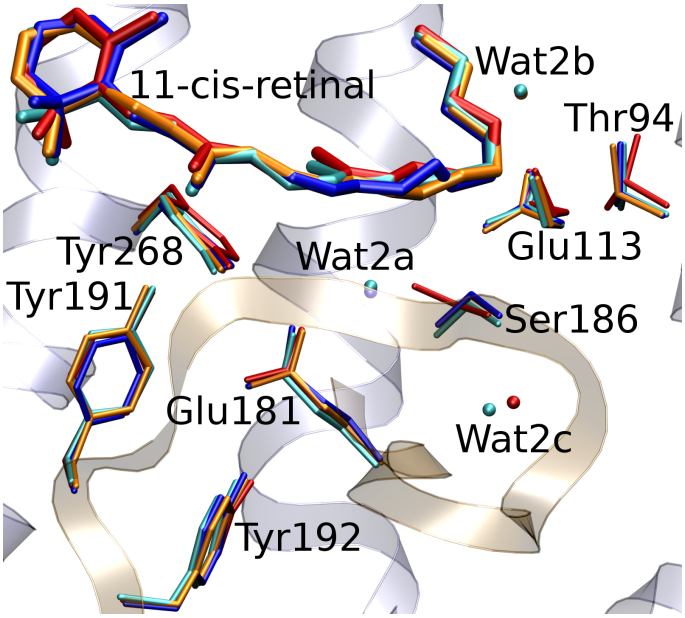


Figure 14: Comparison of X-ray structures 1F88 (red), 1HZX (blue), 1L9H (orange), and 1U19 (cyan).
Wat2a connects Glu181 and Ser186, Wat2b is close to the PSB and the counterion Glu113 and Wat2c is located on the extracellular side of Ser186.

Table 13: Details of published X-ray structures.

PDB-code	year	resolution	waters in active site	total water	color (Fig. 14)
1F88[14]	2000	2.8 Å	Wat2c	16	red
1HZX[15]	2001	2.8 Å	Wat2a	12	blue
1L9H[16]	2002	2.6 Å	Wat2a, Wat2b	13	orange
1U19[52]	2004	2.2 Å	Wat2a, Wat2b, Wat2c	44	cyan
1GZM[53]	2004	2.65 Å	Wat2a, Wat2b, Wat2c	20	—
2J4Y[54]	2007	3.4 Å	—	no water	—

Figure 14 shows an overlay of the available Rh crystal structures: 1F88,[14] 1HZX,[15] 1L9H,[16] and 1U19.[52] Two more structures (not shown) have been published (1GZM[53] and 2J4Y[54]) which are very similar to 1U19, except for the number of resolved water molecules. Table 13 shows some data of the crystal structures. The most recent structure (1U19) resolves three water molecules in the active site, labeled Wat2a, Wat2b, and Wat2c. They are part of an extended HBN on the extracellular side of the chromophore.

The structures have very similar conformations of the residues in the active site except for Ser186 (Figure 15). In all but the 1F88 structure, the Ser186 C_β–O bond is oriented perpendicular to retinal, pointing towards the extracellular side, and is in hydrogen-bonding distance to Wat2a (2.9 Å) and Wat2c (3.5 Å). In the 1F88 structure, on the contrary, the C_β–O bond is parallel to retinal and the oxygen atom is displaced by about 2 Å compared to the other structures. Its position is close to the (in 1F88 unresolved) Wat2a and far away from any possible hydrogen bonding partners. A third configuration of Ser186 was proposed in ref 52, which deviates from the X-ray structures: Supported by QM/MM calculations, the Ser186 side chain was completely rotated to form a hydrogen bond with Glu113.

For crystal structures it is important to know that the structure is not measured but calculated from the electron density. An electron density is measured and an initial guess of a structure is positioned into this density. For some residues it can be complicated to build the right conformation of side chains if the electron density is diffuse. E.g. threonine has two isoelectric groups at C_β (CH_3 and OH) that have a fixed configuration because of the stereochemistry but with a rotation of 180° the two groups can also be differently positioned into a diffuse electron density. Another problem are water molecules in the protein structure. The first Rh crystal structure has no Wat2a resolved. Maybe the Ser186 OH group is rotated differently because of the measured electron density of Wat2a.

The atomic positions of the initial guess are optimized with some force field methods that do not need explicit hydrogen atoms (coarse grained models). Protonation states of titratable residues or different protonation pattern of histidines are not included in these calculations and thus the resulting structures have to be handeled with care. Additionally these optimization methods are well suited

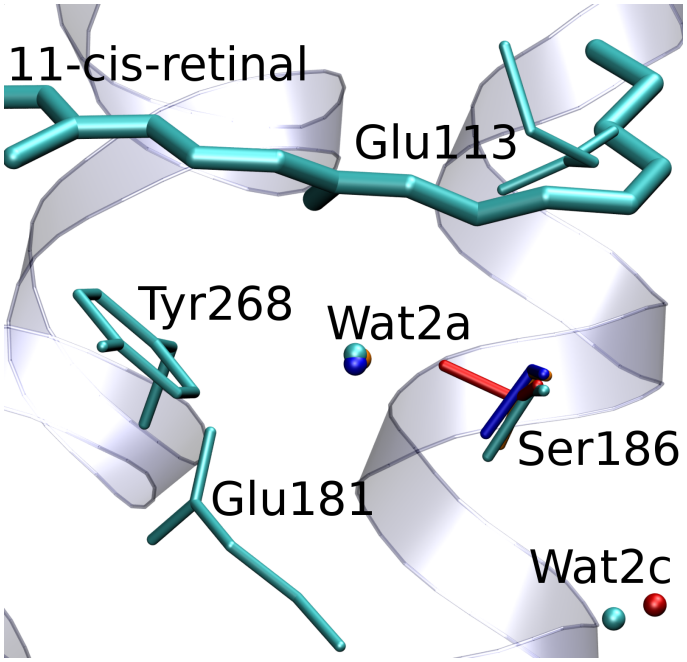


Figure 15: Orientation of Ser186 in the HBN below retinal (for colors see Table 13).

to fit structures of proteins and can have problems with unusual molecules as retinal and also at chemically unusual interactions (PSB–counterion). A simple MM optimization after adding hydrogens is not enough to build a reasonable structure. MD simulations, manual reorientation of some side chains and testing different protonation states should be used to build a good structure for doing serious calculations for research.

4.1.2 Important Role of Glu181

The different states in the photobleaching sequence of rhodopsin have been described in Section 1.3. The important states in this chapter are the ground state and Meta I and II. Glu181 plays an important role in the Meta I to Meta II transition. For these states no or only low resolution structural informations (electron crystallography study[55]) are available but experimental investigations succeeded in building models for this transition. Two different mechanisms are proposed in the literature for this transition that assume different protonation states of Glu181. The first was the so-called counterion-switch mechanism.[11] It assumes a protonated Glu181 in the dark state that transfers its proton to Glu113 during the transition to Meta I, where it becomes the new counterion of the PSB. Structural changes in the binding pocket are proposed to occur bringing Glu181 closer to the PSB. In the dark state the distance is too long (≈ 7 Å). This model is based on two experimental spectroscopic observations: (1) In UV–vis experiments of Rh and site-directed mutants,[11,56,57] the primary counterion of the PSB was identified as Glu113 in the dark state and Glu181 under Meta I-like conditions. Therefore, UV–vis spectra of WT-Rh and E113Q and E181Q mutants were recorded under different pH values before and after photobleaching and trapping the photoproduct in Meta I and Meta II. When one of these two residues

is replaced, the pK_a of the Schiff base drops dramatically in the corresponding state. Hence, Glu113 must be charged in the dark state and Glu181 in Meta I. (2) As discussed in detail in Section 4.1.3, preresonance Raman and UV–vis spectra of WT-Rh and some mutants in the dark state also appear more consistent with a protonated Glu181.

Based on Fourier-transform infrared (FTIR) measurements (also described in Section 4.1.3 in detail), Lüdeke et al.[12] suggested an alternative complex-counterion mechanism. They proposed both Glu113 and Glu181 to be unprotonated in the dark state forming a complex counterion for the PSB. The conformational change towards Meta I brings the PSB closer to Glu181, which then becomes the predominant counterion.

The protonation state of Glu181 is not only important for understanding the activation mechanism of Rh but also for the structure of Rh in the dark state and for all properties depending on this structure (color tuning, PSB accessibility for water and hydroxylamine, etc.).

4.1.3 Experimental Results on Glu181's Protonation State

The question of the protonation state of Glu181 has been investigated using different spectroscopic techniques and theoretical methods, leading to conflicting findings. Some of these studies will be reviewed shortly here. A more detailed description is given in my publication in J. Phys. Chem B. in 2010.[58]

Two-Photon-Excitation

In 1985, Honig et al.[59] suggested several simple point-charge models for the Rh binding site. Their favoured model[60] places one negative charge close to the PSB and another between C_{12} and C_{14} and thus reconciles the observed strong downfield shift of the ^{13}C14 NMR peak in Rh.[61] At that time the structure and especially the infolding of the extracellular loop E 2 close to retinal was not known and thus this second negative charge was not assigned to Glu181.

Also in 1985, Birge et al.[62] measured the two-photon absorption spectrum of 11-*cis*-locked Rh and attempted to reproduce the one and two-photon data in semiempirical CI calculations (INDO/PSDCI). They obtained the best results using a single counterion placed below the polyene between C_{13} and C_{15}. They also claimed that by introducing a second negatively charged group in the binding site, e.g., at the positions suggested by Honig et al.,[60] the energy gap between the S_1 and the one-photon forbidden S_2 state would increase from 0.44 to more than 0.93 eV, whereas their measured gap is ca. 0.25 eV. Hence, they concluded that the binding site must be neutral. This result is very often cited in experimental and theoretical works afterwards and interpreted to provide experimental evidence for a protonated Glu181. But this study has several problems. (1) INDO/PSDCI is a semiempirical method with only single and double excitation. As was shown

by Wanko et al. in 2005[44] multi-reference methods are needed to calculate the excited states of retinal precisely. (2) The measured state at 2.74 eV that is assigned to the S_2 state absorption has been reassigned to a vibrational feature of the S_1 state by Nielsen et al.[63] in 2006. In This study the S_2 state is thought to correspond to an energetically very high band at 3.65 eV (measured by Ebrey and Honig[64]). CASPT2 calculations (with protonated and unprotonated Glu181) by Tomasello et al.[65] find the S_2 state at approximately 3.55 eV consistent with the assignment of Nielsen et al. and with the calculations of Birge et al. with an extra negative charge. (3) As is known from the X-ray structures, Glu181's carboxyl group is located ca. 6 Å away from retinal's backbone (distance from Glu181-C_δ to C_{12}), whereas the point-charge models tested positions at 3 Å separation and neglected screening effects of surrounding polar groups.[59,62]

FTIR

In 1993, Fahmy et al.[66] studied the FTIR difference spectra (dark state–Meta II) [1] of Rh and the site-directed mutants of Asp83 and Glu122. With a site directed mutant the missing bands in the difference spectra can be assigned to the mutated residue. They assigned all difference bands above 1700 cm^{-1} to Asp83 and Glu122, which therefore must be protonated in all intermediate states, and to Glu113, which is protonated in the formation of Meta II. Mutations with Glu181 came later in 2005. In another FTIR study, Nagata et al.[67] see no vibrational changes for the dark state to Batho transition that can be assigned to a protonated Glu181. This argues against a protonated Glu181 in the dark and Batho states of Rh because the strong structural changes caused by the photoisomerization would also induce changes in the vibrational frequencies of a protonated Glu181.

A more recent FTIR study by Lüdeke et al.[12] in 2005 considers also site-directed mutants of Asp83, Glu122, and Glu181. In the difference spectra of dark state to Meta I and dark state to Meta II, peaks were measured at 1767 and 1735/1727 cm^{-1}, which were assigned to protonated Asp83 and Glu122 via D83N and E122Q mutants. The double mutant D83N/E112Q shows a completely flat region above 1700 cm^{-1} in the dark state – Meta I difference spectrum, whereas all peaks in this region from the WT spectrum are unchanged in the E181Q and E181M mutants. Mutation induced changes of protonation states were not discussed in ref 12.

In priciple it can be possible that Asp83 (or Glu122) is unprotonated and Glu181 protonated in the dark state of WT-Rh. The experimental results would then only

[1] FTIR spectra alone give no informations about structural properties because the C=O stretch frequencies of protonated glutamic or aspartic acids are close to the very intense and broad amide I band resulting from all backbone C=O groups. Thus, differences of spectra in dark state and after photoactivation and trapping in Meta II, can give informations on structural changes because the amide I band is not changing much between these states.
In difference spectra, positive and negative peaks show where a peak is in the one or in the other state spectrum.

be explainable if in the D83N (or E122Q) mutant the protonation state of Glu181 is changed. This would suggest a wrong assignment of the missing band to the mutated residue instead of Glu181. But the two bands in the E181Q mutant must stem from protonated Asp83 and Glu122 and it would be a big coincidence that a protonated Glu181 has the same difference band in the D83N (or E122Q) mutant as Asp83 (or Glu122) in the E181Q mutant. This coincidence is very unlikely and additionally the change in protonation state in different mutants is hard to explain because the three residues are far away from each other[2] and are not connected via hydrogen bonds.

In the transition to Meta II, one positive peak appears at 1713 cm^{-1}. This peak must stem from the protonation event of Glu113. As no deprotonation of Glu181 can be detected and Glu181 acts as the principal counterion of the PSB in Meta I, Lüdeke et al. concluded that Glu181 must be unprotonated (charged) in all states.

Raman and UV–vis

Yan et al.[11] compared the preresonance Raman spectra[3] of wild-type (WT) Rh and the E181Q mutant. The modes in the fingerprint, hydrogen out-of-plane (HOOP), and C=C-stretching regions were found to be conserved within a few cm^{-1}. Also the UV–vis absorption of Rh in the dark state is hardly altered by mutation of Glu181.[56,57,68] As the substitution of a negatively charged Glu181 by a neutral Gln is expected to induce significant changes in the vibrational and UV–vis spectra, the authors concluded that Glu181 must be neutral.

But in principle this data can also be interpreted as Glu181 being unprotonated. Therefor, the structural effect of the negative charge in WT-Rh has to be small or similar to the effect of the neutral glutamine. This is possible because Glu181 is located close to the middle of retinal where it was shown that additional charges do not have a large impact on the UV–vis absorption because of a nodal plane in the electrostatic potential of the retinal $S_1 \leftarrow S_0$ excitation.

Reactivity Towards Hydroxylamine

Apart from the spectroscopic data, the analysis of rhodopsin's reactivity with hydroxylamine may yield additional evidence on the matter. Hydroxylamine is a small molecule, that can disrupt the protonated schiff base into a retinaloxime and an amine (Figure 16). The kinetcs of this reaction is dependent on the accessibility of the PSB by hydroxylamine.

[2]Glu122 is located in front of the β ionone ring, 14 Å from Glu181. Asp83 is located 13 Å above retinal, 20 Å from Glu181.

[3]Raman spectra measure the intensity of vibrational frequencies that are invisible in IR spectra due to small or zero change in dipole moment. These frequencies are C–H, C–C, and C=C vibrations. Preresonance Raman spectroscopy gets enhanced band intensities of all vibrations in retinal and not from other protein vibrations.

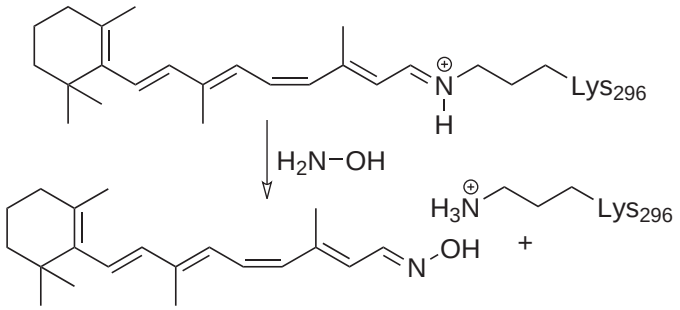


Figure 16: Cleavage of the protonated Schiff base with hydroxylamine.

In all Glu181 mutants (except E181D), the PSB shows a drastic increase in the hydroxylamine reactivity[57] with respect to WT Rh. The accessability of hydroxylamine to the Schiff base depends on the stability of the hydrogen bonded network (HBN) that connects the loop E2 residues (including Glu181) with the remainder of the binding site (Figure 17), as the penetration of hydroxylamine requires the transient breaking of several hydrogen bonds of the HBN. Lewis et al.[69] studied the effect of Glu181 mutation on the early photointermediates. They found a reduced stability of the Batho intermediate for mutants other than E181D, indicating a less stable HBN in the binding pocket.

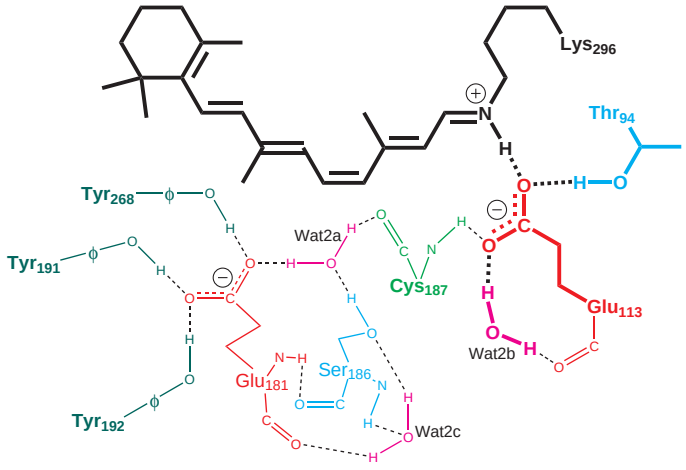


Figure 17: Possible structure of a stable HBN with a charged Glu181, consisting of Glu181, Ser186, Cys187, Tyr191, Tyr192, Tyr268 (helix VI), and Wat2a, Wat2c.

The results indicate an instable HBN for all Glu181 mutants except E181D. Thus, Glu181 and Asp181 are thought to be charged and give a very stable HBN because even the E181Q mutant, that can also build an HBN with its NH_2 group, shows an increased hydroxylamine reactivity.

Janz et al.[70] studied mutants of six amino acids in the HBN. All of them have a drastic effect on the energetics of Schiff base hydrolysis. This is attributed to a weakening of the HBN. Four of the mutants (T94I, S186A, Y192F, Y268F) clearly have a smaller HBN because of a missing hydroxyl group. Also for the counterion mutant E113Q, this observed effect is expected. For the E181Q mutant, this result is surprising and indicates a strong disturbance of the HBN in this mutant, too.

NMR

With NMR measurements alone, no conclusions on the Glu181 protonation state can be drawn. Only in combination with calculations arguments for or against a model can arise.

Very recently, during preparation of the manuscript on this topic (Ref 58), the first ^{13}C chemical shift measurement for the E181Q mutant was published by Ahuja et al.[71] They probed only the C_{12} position, for which they reported an upfield shift of 2.4 ppm compared to WT-Rh.

4.1.4 Computational Results on Glu181's Protonation State

Previous quantum-chemical studies on Rh investigated optical and other spectroscopic properties with Glu181 modeled protonated,[48,51,72–83] and several studies tested both protonation states,[65,84–90] but came to different conclusions. They are reviewed briefly in the following. See Ref. 58 for a more detailed description.

UV–vis

In an early QM/MM study, Schreiber et al.[85] calculated the effect of a protonated or unprotonated Glu181 on the excitation energy of a reduced model chromophore *in vacuo*. They obtained a surprisingly small shift (0.12 eV), and explained this with the position of Glu181 close to the center of the chromophore. They implied that the argument of Yan et al.[57] for a protonated Glu181 (small mutation effect) does not hold.

In classical MD simulations,[86,87] Martinez-Mayorga et al. simulated the Batho to Meta I transition, starting from either a protonated or an unprotonated Glu181. In the latter case, the predominant counterion spontaneously changes from Glu113 to Glu181 during 1000 ns of unconstrained MD simulation. This corroborates the model suggested by Lüdeke et al., as it shows that such a switch is possible without any proton transfer. Grossfield et al.[87] reanalyzed these trajectories and found in both cases an increase in the number of water molecules inside the protein during Meta I formation, consistent with ^{1}H MAS NMR results. Therefore, the authors consider the complex-counterion model with a charged Glu181 in dark and Meta I states to be more likely but do not exclude the proton-transfer-triggered counterion switch.

In a recent QM/MM study, Hall et al.[88] considered both protonation states and the E181Q mutant. They showed that the $S_1 \leftarrow S_0$ and $S_2 \leftarrow S_0$ excitation energies are very similar for models with protonated (2.38 eV) and charged (2.44 eV) Glu181 and also the E181Q mutant (2.38 eV) gives a very similar result. They conclude that it is not possible to decide on the protonation state of Glu181 simply from experimental results as Birge et al. did.

Very recently, Sekharan and Buss[91] tested the effect of the Glu181 protonation state on the CASPT2 excitation energy using a gas-phase model (full chromophore, Glu131, and Glu181 side chain). This work has several problems:

1. Their WT-Rh structure is the same as in Ref. 51 produced with a protonated Glu181. This optimized structure is based on the structural model produced in Ref.[52] with a changed orientation of Ser186.

2. They did not consider any structural changes in the HBN in building the model with a charged Glu181.

3. They calculated the external charges of the binding pocket in a DFT calculation with a natural population analysis. They made an error in taking the charges from a calculation of the binding pocket atoms with the chromophore. The sum of MM charges surrounding the chromophore is therefor not Zero.

4. They did not report any details of their mutant model and its HBN.

5. Their main argument is the difference of the S_0 and S_1 dipole moments $\Delta|\mu|$ that is said to be measured as 12.0 ± 2.0 D by Mathies et al.[92] But in this work the correct quantity is $|\Delta\mu|$. This is a large difference if the two dipole moments of the S_0 and S_1 states are not parallel and this is the case for retinal. Additionally the calculation of a dipole momnt of a charged residue is depentant on the choice of the origin.

Thus, the results of[91] have to be taken with a lot of care. They found a small shift of only 0.09 eV between the protonated and unprotonated model. Only for the protonated model and for the E181Q mutant, they electrostatically included the binding pocket charges and obtained 2.47 eV for the WT and a bathochromic mutation shift of 0.02 eV, consistent with experiment.

In the recent CASPT2 study of Tomasello et al.,[65] the entire protein is electrostatically included in terms of AMBER ff99 point charges. In difference to Sekharan et al., the chromophore is optimized inside the protein using CASSCF instead of SCC-DFTB, and the opsin coordinates are kept fixed to the values in the X-ray structure (1U19). They obtain 2.32 (2.50) eV, assuming a protonated (charged) Glu181. Hence, they argue that a charged Glu181 model is more consistent with experiment.

A common problem in the mentioned computational studies is that the (de)protonation or mutation of Glu181 requires changes in the HBN of the binding site. The HBN, again, is far from obvious and depends on the number of assumed water molecules and on the orientation of the Ser186 and Thr94 side chains, which is not certain from the X-ray structures. The different possibilities for the HBN and their influence on the simulated spectral properties have never been discussed.

Both works by Sekharan et al. and Tomasello et al. show that the $S_2 \leftarrow S_0$ excitation energy is little effected by a change in the protonation state of Glu181.

But the absolute excitation energies are far off from each other, because of different optimization methods and because of a larger QM region and no external charges in Ref. 91 (see Table 14).

Table 14: CASPT2 results from Ref. 91 and 65 (eV). Experimental results are 2.50 eV for $\Delta E_{S_1 \leftarrow S_0}$ and 2.74 or 3.65 eV (measured by Birge et al.[62] or Ebrey and Honig[64])

		Glu181H	Glu181$^{\ominus}$
Sekharan et al.[91]	$\Delta E_{S_1 \leftarrow S_0}$	**2.50**	2.59
	$\Delta E_{S_2 \leftarrow S_0}$	2.92	2.94
Tomasello et al.[65]	$\Delta E_{S_1 \leftarrow S_0}$	2.32	**2.50**
	$\Delta E_{S_2 \leftarrow S_0}$	3.54	3.56

NMR

Also NMR chemical shifts were used to determine the position of a single counterion by matching theoretical with experimental data.[93–95] The resulting position is closer to C_{14} than to the PSB, suggesting the idea of a weak or delocalized counterion. In regard of the first X-ray structures, Verhoeven et al.[96] remark that the structural models inferred from chemical shift constraints locate the counterion somewhat farther away from the PSB than indicated by the X-ray structures. This discrepancy was interpreted in terms of a complex counterion that may involve a charged Glu181 or Tyr268 residue.[97,98]

Martinez-Mayorga et al.[86,87] reported a better agreement between simulated and measured ^{2}H-NMR spectra for the Meta I state[99] obtained with the charged Glu181 in their classical MD simulations.

Hall et al.[88] calculated ^{1}H and ^{13}C chemical shifts of their models with protonated and unprotonated Glu181 and the E181Q mutant and compared the results with the experimental shifts of WT-Rh.[100] They found slightly better agreement with experiment when assuming a protonated rather than a charged Glu181. However, it remains unclear whether the deviations between the models are significant or depend on the computational approach.

Röhrig and Sebastiani[89] computed NMR chemical shifts comparing both protonation states as well. They also found differences between the models to be small compared to the deviations from experiment and concluded that ^{13}C and ^{1}H shifts are not suited for the decision of the protonation state.

Stability of the HBN

The increased hydroxylamine reactivity (see page 66) can not be calculated directly. Only the stability of the HBN on the extracellular side (below) retinal can be calculated from MD simulations.

A difference in HBN stability for Glu181 mutants was found in a theoretical MD study by Röhrig et al.[84]They performed classical MD simulations of Rh in an explicit solvent and membrane environment Assuming a protonated Glu181, they experienced a disruption of the HBN after 630 ps. Starting from an unprotonated Glu181, their HBN was stable during a 7 ns simulation. However, another classical MD simulation over 1.5 μs did not report such instabilities in neither case.[86,87]

4.2 Approach to Solve the Problem of Glu181 Protonation State

The purpose of this study is to eliminate most of the aforementioned uncertainties of the computational studies, simulate the entire range of experimental data that may distinguish the two Glu181 protonation states, and identify additional properties that may lead to evidence in future experiments. First, consistent models for both protonation states and two mutants of Glu181 are constructed and their geometries are compared with the available crystal structures. Protein relaxation is simulated using QM/MM MD simulations and possible side chain orientations are evaluated that are unclear from the X-ray structures (Ser186). IR, Raman, NMR, and UV–vis spectra for all Rh models are computed. The differences between theoretical models and experimental data are analyzed and, where necessary, their dependence on the computational approach is discussed. To obtain a broader basis for this analysis and estimate the sensitivity of the methodological approach, it is necessary to perform the same calculations on similar systems. Two archaeal rhodopsins, pharaonis phoborhodopsin (psRII) and bacteriorhodopsin (bR), are included in this study. In particular for bR, experimental data (two-photon absorbtion, NMR) is available that is valuable to understand the significance of the theoretical data. Based on this critical analysis, previous interpretations of the experimental data with regard of the protonation state of Glu181 are re-evaluated.

4.2.1 System Setup and Methods

The system setup is the standard setup described in Section 2.2.1.1 with the 3rd order extension in DFTB in combination with CHARMM for geometry optimizations, vibrational analysis, MD simulations and free energy calculations.

The standard QM region for all calculations in this chapter is QM4 (Lys296, Glu113, Thr94, Wat2b). Glu181 is treated classically. Only for some calculations bigger QM regions are used, which are called QM5_x and are comprised of QM4 and the amino acid with the sequence number x.

C=O Vibrations (IR)

Calculations of C=O vibrations are performed with QM5_x regions that are comprised of the QM4 region and the side chain of the protonated glutamic or aspartic acid x to be calculated. The structure is optimized with this larger QM region and the VIBRAN module of the CHARMM program package[33] is used to calculate the vibration energies.

Excitation Energies (UV–vis)

For excied states calculations the QM region of DFTB/CHARMM QM/MM-optimized geometries is re-optimized at the DFT/CHARMM level of theory. For retinal model chromophores, the PBE0 functional accurately reproduces the geometry of CASPT2 benchmark calculations,[101,102] in particular the BLA, which has a strong impact on the excitation energy.[44] The re-optimization of the QM region was carried out with fixed MM atoms, using the PBE0 hybrid functional[31,32] and the TZVP basis.[47] The turbomole package[103] was used for optimization, and VdW energy and forces were established via scripts and the "-modus" option. The same QM/MM coupling and linking scheme was used as in the DFTB/CHARMM optimizations.

C=C and C=N Vibrations (Raman)

QM/MM Raman spectra (vibrational frequencies and intensities) were calculated by the group of Maria Andrea Mroginski at TU Berlin for our publication. They use the CHARMM27 force field[34] and DFT (B3LYP/6-31G*). The atoms in the QM region (QM4) were re-optimized with the program package CHEMSHELL.[104,105] The resulting force fields were then scaled with a global set of scaling factors in order to correct systematic errors associated with the DFT method and the harmonic approximation.[106,107] The scaling procedure, the normal-mode analysis, and the calculation of the Raman intensities of protein cofactors by means of the finite-electric-field method were previously described in detail.[106,108]

NMR Chemical Shifts

NMR chemical shifts were alculated by Marius Wanko (group of Marcus Elstner). Geometries are re-optimized at the DFT/CHARMM27 level (QM1 region,

with PBE0 functional and TZVP basis set). For NMR chemical shift calculations, the PBE0 functional is employed, which is advertised for NMR calculations.[109] The SV(P) basis set is used. Test calculations on the retinal chromophore with triple-ζ and IGLO basis sets yield very similar relative shifts but systematically overestimate the average shift between sp2- and sp3-hybridized ^{13}C atoms. The Gauge-Invariant Atomic Orbital (GIAO)[110] procedure is used, as implemented in version B.04 of the Gaussian03 program.[111] The NoFFM option and an "UltraFine" grid for integral evaluation are used (see B.05 release notes). The QM1 region and CHARMM27 charges were used. Surprisingly, the ^{13}C shifts showed little change and were not improved by extending the QM region. ^{1}H shifts, in contrast, were substantially improved when including groups with steric contact to the chromophore (see Ref. 58). All chemical shifts are referenced to a tetramethylsilane (TMS) calculation with the same method and basis set.

Free Energy Profile

The free energy profile for the Ser186 rotation was computed by Prasad Phatak (group of Marcus Elstner). 3rd-order DFTB/CHARMM27 with holonomic internal coordinate constraints[112] and thermodynamic perturbation theory (TPT),[113–115] as implemented in CHARMM is used. The free energy profile is computed as a function of the C–C_α–C_β–O_γ dihedral angle ξ by constraining it to fixed values from $-66\,^\circ$, $-54\,^\circ$, $-42\,^\circ$, ... to $+66\,^\circ$. The Newtonian equations of motions were solved using the Verlet algorithm.[116] For each simulation, the system is gradually heated to 300 K and equilibrated for 100 ps prior to 200 ps of production run. Thus, more than 2.4 ns of trajectories were used for free-energy calculations. QM4 and QM5_186 regions were used in these calculations.

Protonation State of Glu122 and Asp83

Asp83 and Glu122 are assumed to be protonated because of the assignment by Ref. 12. This is validated by the results of C=O vibration calculations (see Section 4.3.3) and calculation of excitation energies with protonated and unprotonated Asp83 and Glu122 (see Appendix D). A charged Glu122 can be ruled out because of the resulting very small excitation energy of the model with unprotonated Glu122 ($\Delta\Delta E \approx 0.7$ eV relative to a model with protonated Glu122). The small experimental shift of the E122L mutant ($+0.02$ eV[117]) is well reproduced by models with protonated Glu122, regardless of the Glu181 protonation state, whereas otherwise mutation shifts larger than 0.6 eV arise. It is further shown that the protonation state of Asp83 has a very small effect on the absorption energy (< 0.03 eV), ΔBLA < 0.001 Å), and NMR chemical shifts (see Appendix D for details). In particular, none of the trends discussed in this work is affected by the protonation state of Asp83. Based on these findings and the FTIR experiments and results (see Section 4.3.3), Asp83 and Glu122 are assumed to be protonated in the following.

Thr94 Orientation

The orientation of the Thr94 side chain in all models differs from the orientation in the X-ray structures for two reasons: (1) The rotated orientation is energetically favored by 8–10 kcal/mol, and (2) the shift in experimental excitation energy from WT-Rh to the T94I mutant is only reproduced with the rotated orientation (see Section 4.3.1).

4.2.2 Structural Models

Starting from the best-resolved structure, 1U19, which has also most water molecules resolved, independent models of WT-Rh, with protonated and unprotonated Glu181, and the mutants E181Q and E181A were prepared. In all cases, the same procedure was applied and alternative conformations were tested, including the two different suggested orientations of the Ser186 side chain (as in X-ray structures, and rotated orientation as in the QM/MM optimized structure of Ref. 52).

An QM/MM structure optimization is started to prevent artificial interactions due to the addition of hydrogen atoms. For all protonation states and mutants an independent charge-scaling calculation was performed. An MD simulation is started with these structures and after heating and equilibration thermally quenched and optimized to give the final structures. Based on these structures, the orientation of Ser186 and the HBN is manually changed and again optimized. The quality of these changed structures is checked with MD simulations (stability) and with the absolute QM/MM energy (relative to the final structures).

4.2.2.1 WT-Rh Models

WT-Rh Models with Protonated Glu181

In the case of a protonated Glu181, different HBNs can be formed in the active site. The starting point is a model described by Okada et al.,[52] in which Ser186 has a hydrogen bond to Glu113 (rotated orientation). In this model (denoted in this work as PROT1), the added hydrogen at Glu181 points to Wat2a (Figure 18, left). Several other structural models of the active site with different protonation patterns of Glu181 and the two orientations of Ser186 are explored, which are discussed in Appendix E.1. All structures are optimized and the resulting structures are compared in terms of the QM/MM total energy. Based on this criterion, the model of Okada et al.[52] turns out to be the most stable one (PROT1, Figure 18 left). A second model (PROT1b), however, is only 1.5 kcal/mol higher in energy than PROT1 and features the X-ray orientation of Ser186 (Figure 18 right). All other structural models are significantly higher in energy.

The free-energy difference between the two Ser186 orientations is computed using thermodynamic perturbation theory (Figure 19). Depending on the size of

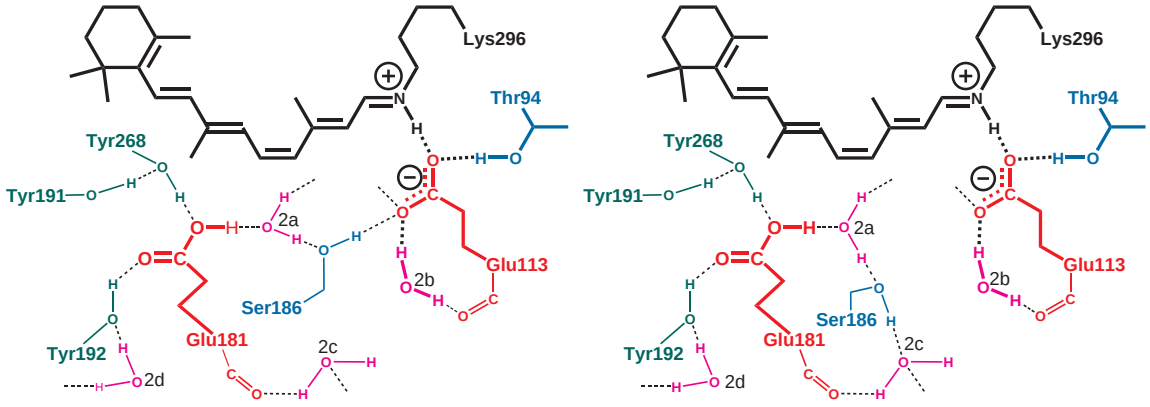


Figure 18: Models with protonated Glu181: PROT1 (left) is only 1.5 kcal/mol higher in energy than PROT1b (right).

the employed QM region (QM4 or QM5_186), a free energy difference of merely 0.1–0.8 kcal/mol is obtained. Considering the approximate nature of the used QM/MM methodology, none of the two models can be excluded a priori based on this small energy difference, and both models will be considered in the following investigations.

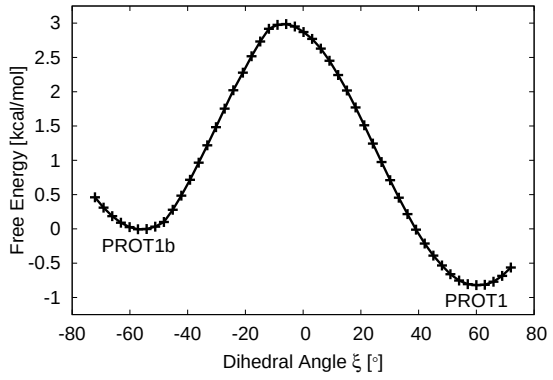


Figure 19: Free-energy profile for the rotation of the Ser186 OH group around the C_α–C_β bond.

WT-Rh Models with Unprotonated Glu181

Three different orientations of Ser186 with unprotonated Glu181 are tested (see Appendix E.2). In all of them Wat2a has a hydrogen bond to Glu181. Two of the models are significantly higher in energy and are therefore not further discussed in the following. In the energetically best model, Ser186 has the X-ray orientation, like in model PROT1b. This structure, which is stable during a QM/MM MD of 2.5 ns, is called UNPROT1 in the following and is shown in Figure 20. Note that structural details can depend sensitively on the size of the QM region. E.g., the orientation of Tyr191, forming a hydrogen bond with Tyr268, is only stable using large QM regions (QM4, QM5_181). With smaller QM regions, Tyr191 rotates to form a hydrogen bond with Glu181 (see Appendix E.2).

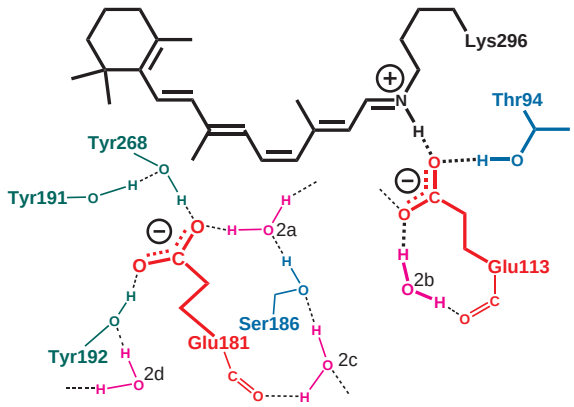


Figure 20: Model with unprotonated Glu181 (UNPROT1).

Summary of WT-Rh Models

The properties of three structures of WT-Rh are investigated in more detail: For unprotonated Glu181, one model (UNPROT1) is clearly identified as the most stable one, whereas for the protonated case, two models (PROT1 and PROT1b) are found that are close in total as well as in free energy. Note, that PROT1 and PROT1b differ only in the Ser186 orientation. All three models are quite similar to the X-ray structure 1U19 (see Figure 21). The mean absolute deviation from the X-ray structure (≈ 0.5 Å) is considerably smaller than the nominal resolution of the latter (1.8 Å). Minor differences in the three models occur in the region of the β-ionone ring and close to Ser186 (Wat2a, Cys187, Phe293). All models have a large displacement at residue Thr94, which is rotated relative to the X-ray structure to form a hydrogen bond to the counterion Glu113.[80]

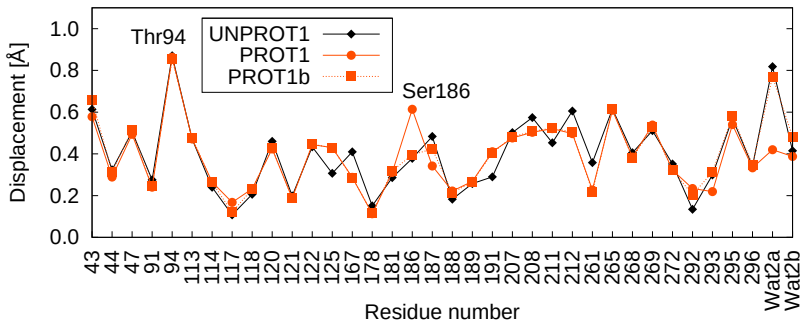


Figure 21: Displacement of non-hydrogen atoms from the 1U19 X-ray structure positions (averaged for every residue). Only residues of the binding pocket (within 5 Å of the chromophore) are shown.

The models also feature very similar chromophore geometries with a similar BLA along the conjugated chain (0.078, 0.076, and 0.081 Å, for PROT1, PROT1b, and UNPROT1, respectively). As the BLA correlates linearly with the absorption energy,[43,48,118] the excitation energies of the three models can be expected to be similar.

The increased BLA in the UNPROT1 model is more pronounced towards the

β-ionone ring, but is continuous over the entire conjugated chain (Figure 22). This is in variance with the unprotonated Glu181 model of Hall et al.,[88] which also shows much larger differences in chromophore bond lengths with respect to their protonated Glu181 model. In their unprotonated model, they assume that Wat2a cannot form hydrogen bonds to Glu181, Ser186, and Cys187 at the same time. Presumably, they did not consider a reorientation of the donating hydrogens of Ser186 and Wat2c. As a consequence of the lacking Cys187–Wat2a bond, they report a displacement of Cys187, which may be responsible for the larger differences in the chromophore geometry.

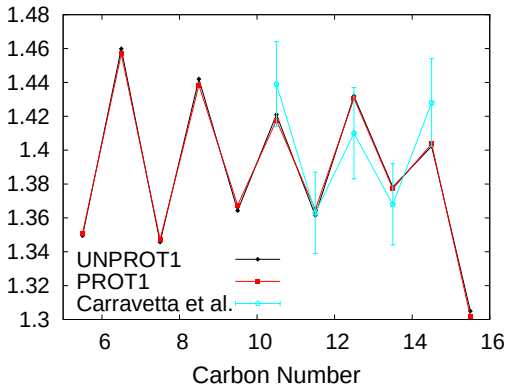


Figure 22: Chromophore bond lengths for models PROT1 and UNPROT1 are very similar and are both within the experimental error of the NMR measurements by Carravetta et al.[119]

4.2.2.2 Model of the E181Q Mutant

Based on the different HBNs in the protonated and unprotonated models, Glu181 is exchanged for glutamine and MD simulations are performed. From these MDs, four different minima (Table 15) are obtained, which are further investigated. The first (GLN1) has the NH_2 group rotated away from the HBN to form stable hydrogen bonds with the backbone oxygen atoms of residues Ile179 and Ile189. This results in a linear HBN (191-268-2a-186). Ser186 has again two possible orientations: the rotated orientation (toward counterion Glu113, GLN1/2/3/4) and the X-ray orientation (toward Wat2c, GLN1b/2b/3b/4b). The second pair of models (GLN2/2b) has a linear HBN (191-268-2a-186) and an NH_2 group which forms hydrogen bonds to Tyr268 and Wat2a. Models GLN3/3b are very similar, with the only difference having a rotated Tyr191, which forms a hydrogen bond toward O_ε of Gln181. Models GLN4 and GLN4b are very different, forming a cyclic HBN (181-268-191-181) and a short linear one (181-2a-186).

These eight models are set up equivalently to obtain comparable energies. Table 16 provides energies of all models relative to the most stable one (GLN1), calculated with QM4 region. MD simulations are carried out to test the stability of the structure. Models GLN2b/3b/4b rapidly convert into the structure of GLN1b. Model GLN2 converts into GLN1, and GLN3 and 4 are in equilibrium with each other.

The model GLN1 is the most stable one and is further investigated in the following.

Table 15: Eight E181Q mutant models and their hydrogen-bonding patterns.

rotated orientation of Ser186	X-ray orientation of Ser186
GLN1 (191-268-2a-186-113) (181-179/189)	**GLN1b** (191-268-2a-186-wat2c) (181-179/189)
GLN2 (191-268-2a-186-113) (181-268/wat2a)	**GLN2b** (191-268-2a-186-wat2c) (181-268/wat2a)
GLN3 (191-181)(181-268-2a-186-113) (181-wat2a)	**GLN3b** (191-181)(181-268-2a-186-wat2c) (181-wat2a)
GLN4 (181-268-191-181) (181-2a-186-113)	**GLN4b** (181-268-191-181) (181-2a-186-wat2c)

Table 16: Energies of E181Q models relative to GLN1 (QM4, kcal/mol).

n	GLN*n*	GLN*n*b
1	0.0	4.5
2	4.9	8.6
3	8.5	10.2
4	11.6	14.3

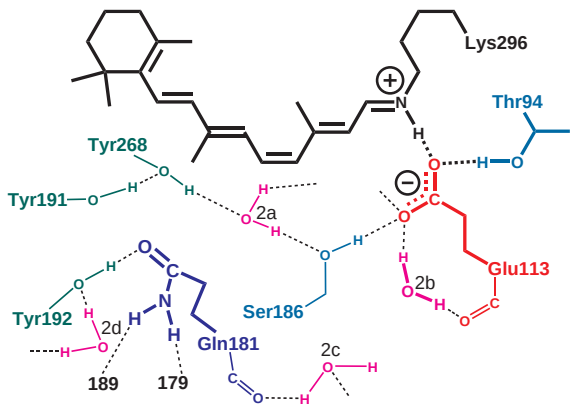


Figure 23: Mutant E181Q model (GLN1).

4.2.2.3 Model of the E181A Mutant

Mutating Glu181 to Ala181 results in the model ALA1, which forms the same HBN as GLN1. The missing side chain of Glu181 leaves a large cavity in ALA1, in which one (model ALA2) or two (model ALA3) water molecules can easily be accommodated (Figure 24). The first water (Wat2e) is strongly bound to the backbone oxygen atoms of residues Ile179 and Ile189, like the NH_2 group in GLN1, and to the O–H group of Tyr192. The second water molecule (Wat2f) is accommodated between the first one (Wat2e), Wat2a, and Tyr268, i.e., takes the position of the Glu181 OH group in the PROT1 and PROT1b models. Both models ALA2 and ALA3 will be considered in the following sections.

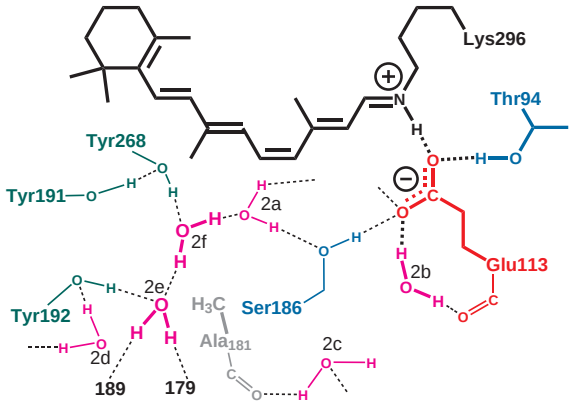


Figure 24: E181A mutant model ALA3 with two added water molecules. The ALA2 model lacks the central water molecule (Wat2f) and the ALA1 model also lacks Wat2e.

4.2.2.4 Stability of the HBN

The reactivity towards hydroxylamine has been related to the flexibility of the HBN by Janz et al.[70] It seems to be intuitive that the HBN of The UNPROT1 model is more stable because of the additional charge. This assumption is confirmed by performing MD simulations (Section 4.3.2. Therefore the strong increase in hydroxylamine reactivity between WT Rh and the E181Q and E181A mutants could be interpreted as another hint for an unprotonated Glu181. Clearly, the calculation of this reaction is beyond current computational possibilities.

4.3 Solution of the Problem of Glu181 Protonation State (Results and Discussion)

4.3.1 Thr94 Orientation

The orientation of the Thr94 side chain in all models differs from the orientation in the X-ray structures for two reasons. 1) The rotated orientation is energetically

favored by 8–10 kcal/mol (see Table 17) and 2) the shift in experimental excitation energy from WT-Rh to the T94I mutant is only reproduced with the rotated orientation (see Table 18). The setup of the models with Thr94 in X-ray orientation is the same as for all other models. In these models the O–H group of Thr94 has a hydrogen bond to the backbone oxygen atom of of Cys185.

Table 17: QM/MM energies (SCC-DFTB/CHARMM, QM4, kcal/mol) of models with different orientations of the Thr94 side chain relative to the rotated orientation used in this work.

Model	ΔE
PROT1	+7.9
UNPROT1	+10.2
GLN1	+8.5

Mutating Thr94 into an isoleucine gives several models which are evaluated energetically and with MD simulations. Two Ile94 orientations are very close in energy (within 2 kcal/mol) and are similar in excitation energy (within 0.05 eV). Both models were chosen to calculate the shifts in excitation energy in Table 18. The experimental shift of +0.11 eV is only reproduced by the energetically lower models with a rotated Thr94 sidechain (PROT1 and UNPROT1).

Table 18: Excitation energies (SORCI, QM4 with CHARMM charges[a], SCC-DFTB/CHARMM optimized, eV) of the used models with different Th94 orientations and shifts to the T94I mutants.

Model	$\Delta E_{S_1 \leftarrow S_0}$	$\Delta\Delta E^{b}_{S_1 \leftarrow S_0}$
T94I (Glu181 prot.)	2.46...2.51	—
PROT1 (T94 rotated)	2.43	+0.03...+0.08
PROT1 (T94-X-ray)	2.53	−0.07...−0.02
T94I (Glu181 unprot.)	2.62...2.66	—
UNPROT1 (T94 rotated)	2.54	+0.08...+0.12
UNPROT1 (T94-X-ray)	2.64	−0.02...+0.02

[a] The use of the polarization model polar.h should not be that important in this case because Thr94 is part of the QM region and the MM region is not changed.
[b] Shifts are from the respective model to the T94I mutant with the same protonation state of Glu181. The experimental shift is +0.11 eV.[70]

4.3.2 Stability of the HBN

Röhrig et al.[84] reported that with a protonated Glu181, the HBN is disrupted after 630 ps of classical MD simulation, whereas it is stable in a 7 ns simulation

when modelling with an unprotonated Glu181. For the PROT1 and PROT1b models 2.5 ns of QM/MM MD simulation (QM1 region) are performed, and do not see such a disruption, consistent with longer simulations in refs 86,87. The sampling times in all these studies are still too short to provide clear statistical evidence for an increased stability of the HBN due to deprotonation of Glu181.

Nonetheless, it is obvious that the HBN is stronger bound in case of an unprotonated Glu181: The hydrogen bonds to Glu181 in the PROT1 and PROT1b models are much longer than in the UNPROT1 model (O–O distances: 2.80 vs 2.60 Å for Tyr268–Glu181 and 2.66 vs 2.62 Å for Wat2a–Glu181, see Figure 26). During the 2.5 ns MD simulations of PROT1 and PROT1b, Wat2a rotates more frequently. Whereas in the case of UNPROT1, the hydrogen bonds are shorter and more stable. In the linear HBN of the GLN1 mutant model, hydrogen bonds are longer because it involves one less residue. In consequence, Wat2a rotates frequently, as in the PROT1 simulation. In the ALA3 model, the HBN is even more flexible because more water molecules are involved(see Figure 25).

In summary, very similar HBNs in the WT-Rh models with protonated Glu181 and in the Glu181 mutants were found with respect to the orientation of Wat2a and Ser186 dipoles. They contain weaker hydrogen bonds than the HBN of the UNPROT1 model. Thus, the drastic change in energetics of Schiff base hydrolysis (flattening of the convex Arrhenius plot of WT-Rh) for the E181Q mutant that is attributed to perturbations in the HBN,[70] cannot be explained with a protonated Glu181. If the measured increase in hydroxylamine reactivity in the E181Q and E181A mutants[57] is indeed related to HBN stability, this is also supported only by the UNPROT1 model but not by models with protonated Glu181.

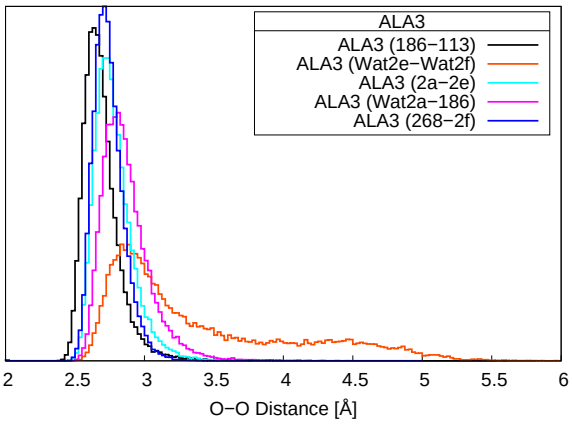
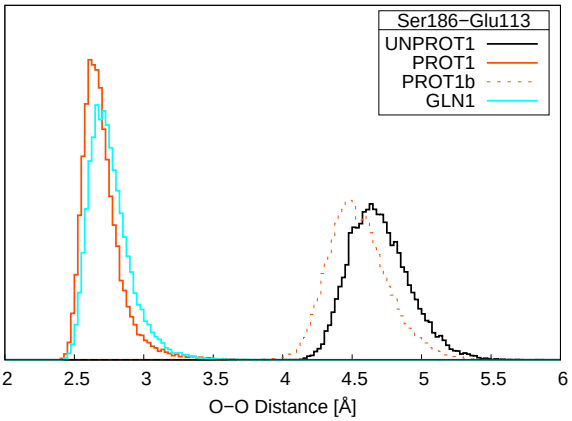


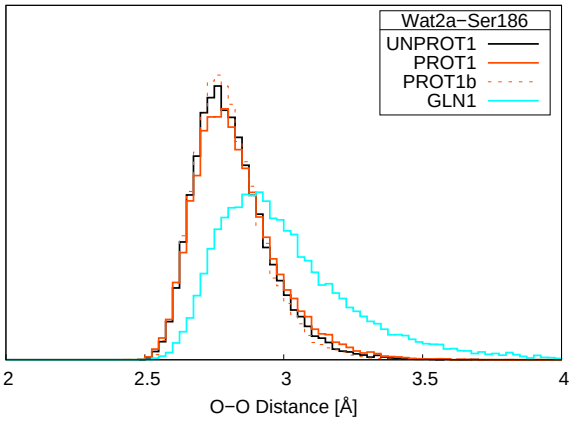
Figure 25: Histogram from 2.5 ns MD simulation of O–O distances in the ALA3 model. The high flexibility of the HBN can be seen at the hydrogen bond between Wat2e and Wat2f, which is partially disrupted.

4.3.3 C=O Vibrations (IR)

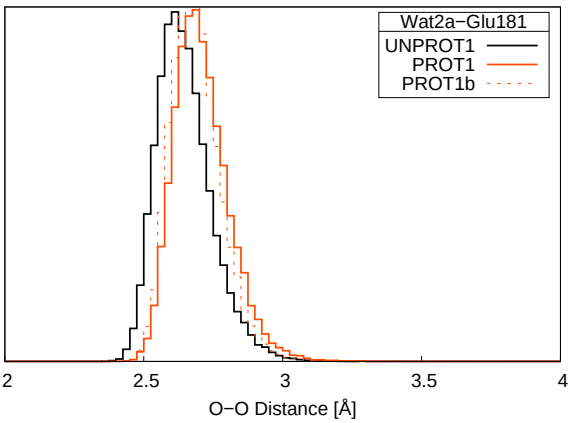
Changes in the protonation state or hydrogen bonding of protonated acids (glutamic or aspartic acids) that occur between different intermediate states can be monitored in FTIR difference spectra in the frequency range clearly above 1700 cm^{-1} (1712–1788 cm^{-1}),[23] where only C=O stretch modes of protonated Glu or Asp residues appear.



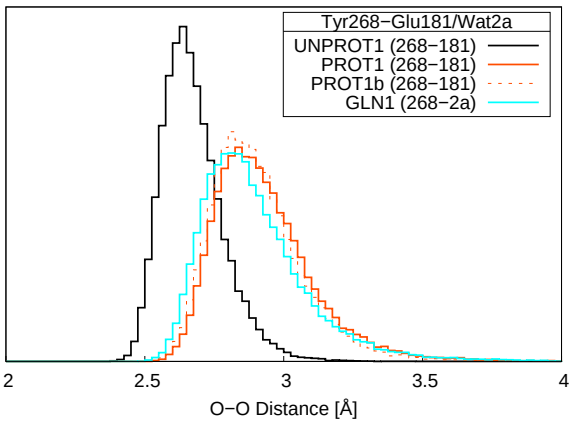
(a) Ser186-Glu113 O–O distance (no hydrogen bond in models PROT1b and UNPROT1).



(b) Wat2a-Ser186 O–O distance (large flexibility in GLN1 model).



(c) Glu181-Wat2a O–O distance (shorter in UNPROT1 model).



(d) Tyr286-Glu181 O–O distance (much shorter in UNPROT1 model).

Figure 26: Histograms of O–O distances in the HBN from 2.5 ns MD simulations.

FTIR studies on Rh[12,66,67] suggest Glu181 to be unprotonated in all states (dark state, Batho, Meta I, and Meta II). This conclusion is due to the absence of any difference bands in the region above 1700 cm^{-1} that can be assigned to a protonated Glu181. As Glu181 is known to be unprotonated in Meta II, there must be a peak corresponding to the deprotonation in the difference spectra. To verify this conclusion and exclude an anomalously shifted C=O stretch frequency, vibrational spectra of the protonated acids (Asp83, Glu122, Glu181) in the PROT1 and PROT1b models are calculated.

Table 19 shows the experimental results from Ref. 12 and the calculated C=O stretch frequencies for protonated Asp83, Glu122, Glu181 (in model PROT1 and 1b) and of Gln181 (in model GLN1). Glu122 produces two bands in the difference FTIR spectra. This can be explained with the orientation of the nearby Cys167, which can have an hydrogen bond with Glu122 or not. With the two possible

orientations of Cys167 two frequencies with a shift of 9 cm^{-1} are obtained, very similar to experiment.

Table 19: C=O vibrational frequencies of protonated carboxylic acids in rhodopsin (cm^{-1}).

acid	calc.[a]	dark[b]	Meta I[b]	Meta II[b]
Asp83	1755	1767	1752	1747
Glu122	1734/1743	1727/1735	1702/1717	1713/1716
Glu181	1729			
Gln181	1691	1707	1707	1707
Glu113	—	—	—	1713

[a] QM/MM (2nd order DFTB/CHARMM) with QM5_X (QM4+181/122/83).

[b] Vibrational frequency assignment from Ref. 12.

Asp83 and Glu122 were assigned neutral by Ref. 12 mainly because of unchanged bands in the E181Q and E181M mutant spectra and one missing band in the D83N and E122Q mutant spectra. Glu122 can be excluded to be protonated in regard of the calculated excitation energies, but the FTIR results would also be explainable with a coupling of the protonation states of Asp83 and Glu181 (see Appendix D). This scenario, however, appears very unlikely because (1) they are far away from each other (17 Å) and not connected via a hydogen-bonded network, Asp83 is screened by two water molecules and has a small impact on the binding pocket. (2) The C=O stretch frequencies of Glu181 and Asp83 would have to be equal and also the shifts to Meta I and Meta II would have to be the same to produce the same difference bands (1767 cm^{-1}) in the D83N and E181Q mutant. The calculated frequencies are clearly seperated by 26 cm^{-1}. The experimental C=O stretch frequencies are well reproduced within few cm^{-1} with the assignment of a protonated Asp83.

For Glu181, PROT1 and PROT1b models predict a C=O stretching frequency of 1729 cm^{-1}. When enlarging the QM region further to include all residues which are hydrogen-bonded to Glu181 (Glu181, Tyr192, Tyr268, and Wat2a), it reduces to 1706 cm^{-1}. Benchmark calculations with small molecules in Refs. 120, 121 and Table 20 show that the method (DFTB/CHARMM) reproduces experimental C=O stretch frequencies in the absence of strong hydrogen bonds within a few cm^{-1}. For systems with strong hydrogen bonds, e.g. the acetic-acid dimer, these frequencies are generally underestimated by about 15 cm^{-1}. Glu181 in the PROT1 and PROT1b models is strongly hydrogen bonded by Tyr192, Tyr268 and Wat2a. Therefore, the FTIR peak of a protonated Glu181 is expected to be at $\approx$ 1720 cm^{-1}. Its absence in the FTIR difference spectra is a strong indication for an unprotonated Glu181.

Lüdeke et al.[12] determined the glutamine C=O stretching frequency of the E181Q mutant at 1707 cm^{-1}. This corresponds to a weakly hydrogen-bonded C=O, as this mode is measured at 1668 – 1687 cm^{-1} in water.[23] The computed C=O frequency of propionamide is 1734 cm^{-1} in vacuo and 1702 cm^{-1} being

Table 20: Benchmark of C=O vibrational frequencies (cm^{-1}).

	DFTB[a]	DFT[b]	exp[c]
acetic acid	1776	1756	1779
acetic acid dimer	1704	1689	1720
propionic acid	1770	1751	1776
propionamide	1734	1704	
propionamide+water	1702	1673	

[a] 2nd order DFTB.
[b] DFT calculation PBE/6-311++g(2d,2p). DFT frequencies are generally too low by about 15 – 30 cm^{-1}.
[c] Vibrational frequencies in argon-matrix from Ref. 122.

hydrogen-bonded to one water molecule (see Table 20). In the GLN1 model, the Gln181 C=O group forms a strong hydrogen bond to Tyr192, which is conserved in all investigated structural models. Consequently, this shifts the C=O stretching frequency down to 1691 cm^{-1}. Therefore, the model cannot account for the unusually high C=O stretching frequency seen in the FTIR experiment. However, it is difficult to imagine any structure, where the Gln residue would exhibit 'gas-phase'-like properties.

4.3.4 C=C and C=N Vibrations (Raman)

The experimental preresonance Raman spectrum[11] exhibits five characteristic bands. Lowest in energy are the HOOP modes at 969 cm^{-1}, the fingerprint modes at 1214, 1237, and 1267 cm^{-1}, which are mixtures of C–C stretching and C–H rocking modes. The broad and intense band at 1547 cm^{-1} is called the ethylenic-stretch band and includes the delocalized stretching of all five C=C bonds. A very small peak at 1655 cm^{-1} can be assigned to the C=N stretch mode.

Yan et al.[11] showed that the WT-Rh vibrational frequencies of the C=C stretch band at 1547 cm^{-1} and of the C=N stretch frequency at 1655 cm^{-1} are only marginally changed in the E181Q mutant.

As reported in Section 4.2.2.1 on page 76, the calculations show that the BLA of the chromophore is quite similar in the models with protonated and with unprotonated Glu181. As the BLA strongly correlates not only with the absorption maximum but also with the C=C stretch frequency, significant changes in these properties upon protonation of Glu181 are not expected (see Section 4.3.5 for UV–vis results). To verify this assumption, Raman spectra were calculated by the group of Maria Andrea Mroginski at TU Berlin.

As shown in Table 21, the calculated spectra of the WT-Rh models with protonated and unprotonated Glu181 are both in agreement with experiment in the range from 1500 to 1700 cm^{-1}. Figure 27 shows that the calculated spectra show little deviation from the spectrum of the E181Q mutant. The C=C stretch

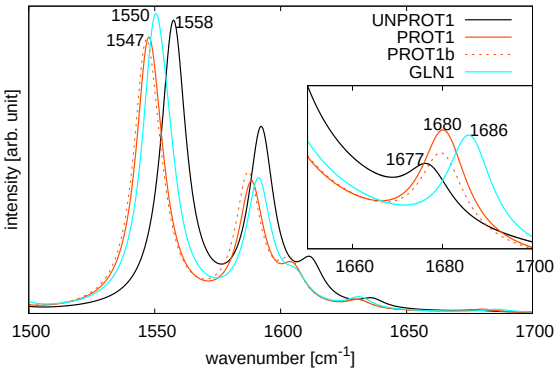


Figure 27: Calculated Raman spectra of WT-Rh models with protonated and unprotonated Glu181 and the E181Q mutant model.

frequencies of WT-Rh are slightly higher than in the E181Q mutant (+4 cm^{-1}) and the C=N stretch frequency is slightly lower (-4 cm^{-1}). The assignment of the vibration at 1677–1686 cm^{-1} is confirmed by the calculation of the potential energy distribution matrix (PED) for these models. It shows that the bands have a strong C=N stretch character coupled with rocking movements of the NH group attached to it. At this point, it is important to emphasize that the variations in the vibrational frequencies found for the three models are smaller than the method's accuracy for the prediction of absolute frequencies.[123]

The shifts from the UNPROT1 model (+8 and -9 cm^{-1}) fit slightly better to the experimental ones than the shifts from the models with protonated Glu181 (-3 and -6 cm^{-1}). This should not be over-interpreted as evidence for an unprotonated Glu181, but our results clearly show that the small shifts found in the experimental spectra of the E181Q mutant cannot be considered indicative for the Glu181 protonation state.

Table 21: Calculated and experimental vibrational frequencies (cm^{-1}).[a]

	GLN1	UNPROT1	PROT1[b]	E181Q[c]	Rh[c]
	1550	*1557*	**1547**		
	1555	**1558**	*1552*		
C=C	1591	1592	1588 (1587)	1543	1547
	1606	*1612*	*1605 (1604)*		
	1631	*1636*	*1631*		
C=N	1686	1677	1680	1659	1655

[a] Numbers in **bold** are the frequencies with highest Raman intensity, numbers in *italics* are very small peaks.
[b] Frequencies for PROT1b are given in parentheses if different from PROT1.
[c] Exp. data from Ref. 11.

Table 22: Excitation energies ΔE (SORCI(man), QM4, eV).

species	model	$\Delta E_{S_1 \leftarrow S_0}$	$\Delta E_{S_2 \leftarrow S_0}$	$\Delta E_{S_2 \leftarrow S_1}$
PSB11	vacuo	2.00	2.37	0.37
	exp[a]	2.03	3.18	1.15
Rh	UNPROT1	2.60	2.97	0.37
	PROT1	2.46	2.88	0.42
	PROT1b	2.48	2.88	0.40
	exp[b]	2.47	2.72	0.25
E181Q	GLN1	2.53	2.90	0.37
	exp[c]	2.44		
E181A	ALA2	2.57	2.95	0.38
	ALA3	2.56	2.91	0.35
	exp[c]	2.48		
bR	see Ref. 43 (QM1)	2.36	2.64	0.28
	exp[d]	2.17	2.62	0.45
psRII	see Ref. 43 (QM1)	2.68	2.79	0.11
	exp[e]	2.50		

PSB11 refers to 11-*cis*-6-s-*cis* dimethyl PSB.
[a] From Ref. 124. [b] From Ref. 57,62.
[c] From Ref. 57. [d] From Refs. 5,125. [e] From Ref. 126.

4.3.5 Excitation Energies (UV–vis)

Some of the calculation in this section were performed by Marius Wanko (group of Marcus Elstner). These are the re-optimizations with PBE0 and the calculations of excitation energies with the polar.h model.

To test the influence of Glu181 mutation and its protonation state on the absorption properties, we performed *ab initio* QM/MM calculations of the S_1 and S_2 excitation energies. The results (Table 22) for the $S_1 \leftarrow S_0$ excitation clearly show that the effect of the protonation state of Glu181 on the absorption maximum is quite limited. With all three WT-Rh models, the opsin shift (from gas phase) and the absolute $S_1 \leftarrow S_0$ excitation energy are in reasonably good agreement with experiment. In the UNPROT1 model, S_1 is blue-shifted by merely 0.14 and 0.12 eV compared to the PROT1 and PROT1b model, respectively. Furthermore, the shift between the two mutants E181Q and E181A is exactly reproduced. Compared with the UNPROT1 model, the absorption is slightly red-shifted in all mutant models (-0.07, -0.03, and -0.02 eV for GLN1, ALA2, and ALA3 models, respectively) very similar to the experimental shifts of -0.03 and $+0.01$ eV for the E181Q and E181A mutants, respectively. In contradiction to experiment, the absorption in all mutant models is blue-shifted compared with the PROT1 and PROT1b models (PROT1: 0.07, 0.11, and 0.12 eV, PROT1b: 0.05, 0.09, and 0.10 eV for GLN1, ALA2, and ALA3, respectively).

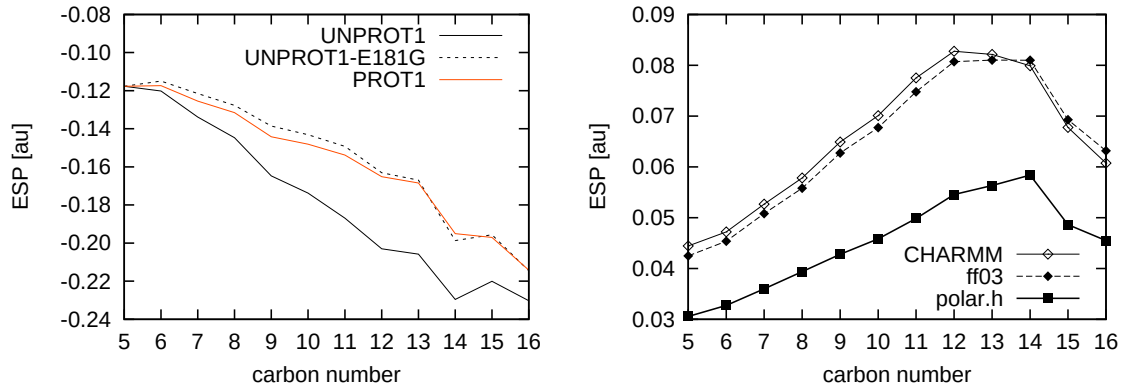


Figure 28: Left: ESP along retinal's conjugated chain, as produced by the CHARMM charges. Right: ESP difference between PROT1 and UNPROT1, as produced by various protein electrostatic models.

The limited effect of protonation state and mutation of Glu181 is obviously due to its distance and position relative to the chromophore. As has been noted before,[69,85] the carboxyl group is located close to the center of the chromophore, where the electrostatic potential (ESP) of the $S_1 \leftarrow S_0$ difference density changes sign.[62] The excitation energy does not respond to a perturbation by a charge monopole in this region, in contrast to that of a dipole moment. This explains, why the mutation shifts for the PROT1 and UNPROT1 behave in such a counterintuitive way. Figure 28 (left) compares the ESPs that are produced by the opsin in the UNPROT1 and PROT1 models along the chromophore backbone. The difference between the two models is dominated by the charge monopole of Glu181 in the UNPROT1 model: when the point charges of the Glu181 side chain are deleted (UNPROT1-E181G in Figure 28 left), the ESP of the PROT1 model is essentially reproduced. Therefore, the qualitative effect of the protonation state is independent from the actual electrostatic representation of the protein. The ESP difference is maximal between C_{12} and C_{14} and decays again towards the PSB. When the CHARMM or AMBER ff03 charges are substituted by the more precise polar.h model, the field of the Glu181 charge monopole is reduced rather homogenously due to protein polarization (Figure 28 right). The net effect on the ESP and excitation energies, however, depends on the actual point-charge model, the QM region, and the opsin as shown in Figure 23.

Table 23 shows the influence of the size of the QM region, the method used for geometry optimization and the electrostatic representation of the protein. When the QM1 region (chromophore only) is extended to include the complex counterion (Glu131, Thr94, and Wat2b for the Rh and Asp85/75, Asp212/201 and three water for the bR/psRII setups) charge exchange between chromophore and the counterion complex causes a hypsochromic shift. In the Rh models, the latter is compensated by the effect of geometry changes due to structure relaxation with the larger QM region. In case of the archaeal rhodopsins, a net hypsochromic shift remains.

Table 23: Excitation energies $\Delta E_{S_1 \leftarrow S_0}$ of different QM/MM setups (SORCI(man), eV).

species	model	QM1 DFTB CHARMM	QM4 DFTB CHARMM	QM4 PBE0 CHARMM	QM4 PBE0 polar.h
Rh	UNPROT1	2.55	2.54	2.67	2.60
	PROT1	2.41	2.43	2.57	2.46
	PROT1b	2.37	2.38	2.51	2.48
E181Q	GLN1	2.44	2.47	2.60	2.53
E181A	ALA2	2.47	2.50	2.61	2.57
	ALA3	2.46	2.51	2.63	2.56
bR	see Ref. 43	2.32	2.41	2.50	2.36
psRII	see Ref. 43	2.56	2.63	2.86	2.68

The shifts in excitation energy (Table 24) that are interesting for the protonation state of Glu181 are only slightly changed by the re-optimized QM regions, and the polarisation model. For the best method the shifts between E181Q and the E181A mutant models is still in excellent agreement with the experimental value and the shifts between WT-Rh and the E181Q mutant models are also in better agreement assuming an unprotonated Glu181.

Table 24: Shifts in excitation energies $\Delta\Delta E_{S_1 \leftarrow S_0}$ of different QM/MM setups, relative to the GLN1 model (SORCI, eV).

species	model	QM1 DFTB CHARMM	QM4 DFTB CHARMM	QM4 PBE0 CHARMM	QM4 PBE0 polar.h
Rh	UNPROT1	+0.11	+0.07	+0.07	+0.07
	PROT1	-0.03	-0.04	-0.03	-0.07
	PROT1b	-0.07	-0.09	-0.09	-0.05
	exp.		+0.03		
E181A	ALA2	+0.03	+0.03	+0.01	+0.04
	ALA3	+0.02	+0.03	+0.03	+0.03
	exp.		+0.04		

The results disclose the difficulties involved in predicting absolute excitation energies of rhodopsins, which depend substantially on the details of the computational approach. It is therefore obvious that the agreement between experimental and theoretical absorption properties for an individual system does not prove the validity of the underlying structural model and assumptions. The rather fortunate agreement of the PROT1 or PROT1b $S_1 \leftarrow S_0$ excitation energy with experiment, e.g., becomes irrelevant when considering the deviations for the Rh mutants and two archaeal rhodopsins in Table 22. In bR and psRII, dispersion

interactions with nearby aromatic residues are expected to red-shift the absorption by ca. 0.07 eV, as is shown in a the dissertation of Marius Wanko.[127] This results in a systematic overestimation of the $S_1 \leftarrow S_0$ excitation energy in all opsin models by ca. 0.1 eV, except PROT1(b). In this perspective, the UNPROT1 results appear more consistent with the remaining data. The overestimation of the opsin shift (with respect to gas phase) by 0.1 eV is a realistic error for the best applicable QM methods, but can also be accounted for by the general structural uncertainties of the X-ray or the QM/MM-optimized geometry.

The calculation of the $S_2 \leftarrow S_0$ excitation energy is more involved and depends strongly on the applied QM method and the underlying geometry. In particular, the S_2 state is strongly affected by the BLA of the polyene chain. Using PBE0-optimized geometries, the gap between S_1 and S_2, as predicted by SORCI, OM2/MRCI, or CASPT2[65,91] in rhodopsins are in reasonable agreement with experiment. Furthermore, the gap hardly changes when Glu181 is deprotonated or mutated (Table 22). In agreement with gas-phase and QM/MM CASPT2 calculations,[65,91] we find that deprotonation of Glu181 slightly decreases the S_1–S_2 gap. In other words, the S_2 state is even less sensitive to the Glu181 protonation state than the S_1 state. Therefore, the arguments of Birge et al.[62,128,129] clearly do not apply to the protonation state of Glu181 in WT Rhodopsin. Their conclusion of a neutral binding site is maybe not due to errors in the computational approach. It appears rather likely, that they simply did not consider the combination of a strong counterion close to the Schiff base and a weakly interacting, charged Glu181 that is separated by 5 – 6 Å from the polyene.

4.3.6 NMR Chemical Shifts

Calculation of NMR chemical shifts were by Marius Wanko (group of Marcus Elstner). Detailed results are given in our publication (Ref. 58). This work includes only the results that are needed to understand the conclusions from these calculations.

Table 25 shows calculated ^{13}C NMR chemical shifts based on the different WT-Rh models, together with results for bR and experimental data. All experimental shifts are from solid-state magic-angle spinning (MAS) experiments with similar sample preparation and agree typically within 0.3 – 0.6 ppm. Therefore, the average experimental value is used as reference. In all calculations, the largest deviation from experiment occurs at the C_{15} atom. The severe underestimation of this downfield shift (ca. 7 ppm) occurs independently of the considered protein, the structural model, the employed point charges, and can be found also in previous B3LYP NMR calculations on the PSB in WT-Rh and in solution.[137] It is idicated that the deviation in the C_{15} shift originates from the QM method rather than the underlying structural model. Hence, C_{15} is excluded from the following considerations. The resulting RMSD to the averaged experimental shifts amounts

Table 25: ^{13}C NMR chemical shifts (ppm) for retinal in Rh and bR.[a]

^{13}C	UNPROT1[b]	PROT1[b]	PROT1b[b]	GLN1[b]	exp.(Rh)[c]	bR[b]	exp.(bR)[d]
1	31.4	31.7	31.6	31.7	34.0	31.9	34.5
2	38.8	38.4	38.4	38.6	40.3	42.9	42.1
3	19.1	18.6	18.8	18.6	20.3	18.3	18.6
4	32.9	33.1	33.0	33.4	34.0	36.6	34.6
5	132.0	136.3	136.8	136.4	130.6	150.7	144.8
6	135.9	136.0	135.9	135.8	137.4	132.4	135.4
7	133.6	138.4	138.9	137.5	132.6	132.0	129.5
8	137.7	136.1	136.0	136.6	139.1	130.1	132.7
9	149.1	152.0	152.5	151.1	148.7	147.8	146.4
10	129.0	127.4	127.3	128.1	127.6	132.1	133.0
11	141.6	141.8	142.0	142.1	141.4	136.9	139.1
12	130.9	127.6	128.2	127.8	132.2	134.8	134.3
13	167.8	164.8	165.7	165.4	168.2	165.2	164.8
14	119.8	117.8	118.2	118.4	121.7	120.1	122.0
15	158.7	158.1	156.9	157.9	165.4	151.6	160.0
16	33.5	33.4	33.3	33.7	30.7	29.8	28.9
17	28.3	28.3	28.4	28.4	26.3	29.3	28.9
18	24.8	25.0	25.0	25.2	21.9	21.5	22.0
19	14.9	14.9	15.0	14.9	13.4	12.9	11.3
20	19.1	18.8	18.9	19.0	16.1	16.3	13.3
RMSD(20)	2.3	3.4	3.6	3.3		2.9	
RMSD(19)	1.8	3.1	3.1	2.9		2.2	

[a] All shifts are relative to TMS; atoms are numbered following the conventional scheme.[130] RMSD wrt. the average experimental value; RMSD(19) excludes C15 (see text).
[b] PBE0/SV(P) NMR calculation.
[c] Average of the exp. data from refs 96,97,131–134,136.
[d] Light-adapted bR from ref 135.

to 1.8 ppm using the UNPROT1 structure, which is similar to the RMSD for the bR results (2.2 ppm), and is merely a factor 2 – 3 larger than the experimental uncertainty. Using the PROT1 and PROT1b structures, the RMSD is clearly larger (3.1 ppm), which is slightly better than the B3LYP/AMBER study of Gascon et al. (RMSD without C_{15} of 3.6 ppm),[137] which also assumed a protonated Glu181. The works of Röhrig and Sebastiani[89] and Hall et al.[88] are the only theoretical NMR studies that compare different WT-Rh models with protonated and charged Glu181. In both cases, the errors in the calculated chemical shifts are much larger than the differences between the two considered models. Hence, no conclusions about the correct protonation state could be drawn from them.

In comparison, the results from the UNPROT1 model agree remarkably well with experiment.

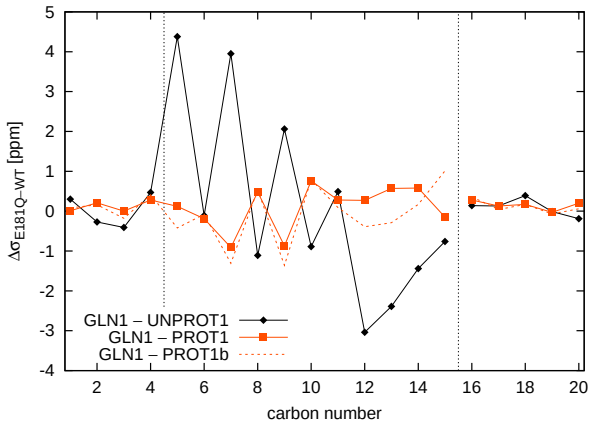


Figure 29: NMR chemical shift differences E181Q mutant (GLN1 model) vs. WT-Rh models.

Finally, the NMR chemical shifts of the GLN1 mutant model are studied. Figure 29 shows that the mutant's chemical shifts agree closely (RMSD 0.4 – 0.6 ppm) with the protonated ones, whereas they considerably deviate from the UNPROT1 ones. In particular, C_5, C_7, and C_9 shifts of the UNPROT1 model are higher by 2.1 – 4.4 ppm, while C_{12} and C_{13} are significantly lower.

In the experimental study of Ahuja et al.[71] the C_{12} position was probed, for which they reported the E181Q utant to be higher by 2.4 ppm compared to WT Rh. This is in good agreement with the result for the UNPROT1 model (3.1 ppm). It is unlikely that this result can be reproduced with any WT-Rh model that assumes a protonated Glu181.

4.4 Summary and Conclusions

In this work, WT-Rh models with protonated, unprotonated, and mutated Glu181 residue were compared. For each scenario, several structural models were constructed and the most realistic one by means of energy and stability in dynamical simulations were selected. The models not only differ in the protonation state of residue 181 but also in the hydrogen-bonding pattern in the HBN, especially in the orientation of the Ser186 side-chain (towards or away from Glu113).

Interestingly, the lowest-energy model for the protonated case, PROT1, is not in agreement with the available crystal structures, as it differs markedly in the position of the Ser186 side chain. A second model, PROT1b, features the same Ser186 orientation as the crystal structures. It is only 1.5 kcal/mol higher in energy than PROT1, the free energy difference is even below 1 kcal/mol. As no QM/MM approach achieves an accuracy of 1 kcal/mol, PROT1b is also regarded as a possible model of the WT-Rh active site. For the unprotonated case, only one plausible model was found, which also agrees well with the crystal structure. All other conformations were significantly higher in energy. The same applies to the models of the E181Q and E181A mutants. For the E181A mutant two models with one or two extra waters are studied that are very similar in all properties.

Despite changes in the HBN, the charge monopole of the unprotonated Glu181 dominates the difference in the ESP along the chromophore when compared to the protonated or mutant models. Due to the position of this charge close to the center of the conjugated chain, protonation has only a small effect on the average electric field in the direction of the chain, i.e., the potential difference between the two ends of the conjugated chain. The latter property, however, is relevant for electrostatically-induced shifts in both the absorption maximum in the UV–vis spectrum and in the C=C stretch vibrational frequencies. In fact, a position of Glu181 slightly closer to the ionone ring would render these properties completely inert to its protonation state. This finding explains the difficulty of theoretical studies to discriminate between the two scenarios.[51,65,84,85,88,89,91]

Already in the structural analysis of the models, virtually no difference in the retinal conformation was found. In particular, the BLA varies by less than 0.005 Å. As the BLA is directly related to the C=C stretch frequency,[43] which in turn correlates with the absorption maximum,[118] no large differences in both, preresonance Raman and optical spectra, are expected.

Indeed, the Raman calculations only show changes of few cm^{-1} in the retinal C=C and C=N stretch frequencies due to protonation or mutation of Glu181. While the small change due to mutation is commonly interpreted as indication for a protonated Glu181,[11] the calculations clearly show that a charged Glu181 leads to the same small shifts upon mutation. In fact, the errors in the mutation shifts are even smaller when referenced to the UNPROT1 value. However, the error is acceptable for all models and care has to be taken not to overestimate the accuracy of the computational methods. Hence, it can be concluded that the Raman vibrational spectra of both models are consistent with experiment.

The results of the excited-state calculations have to be interpreted with some care as well. At first sight, the excitation energy of the PROT1b model appears in perfect match with experiment. However, one should keep in mind that calculations are not at all able to predict excitation energies with an error less than 0.1 eV due to structural uncertainties, intrinsic errors of the QM method, or the QM/MM approximation itself. In the comparison with excitation energies computed for the E181Q and E181A mutants and for the archaeal rhodopsins bR and psRII, a systematic overestimation of the absolute excitation energy by ca. 0.1 eV is found. For the archaeal rhodopsins this shift is rather 0.2 eV due to the neglect of dispersive interactions between chromophore and nearby aromatic side chains, which are negligible in the case of Rh. As a consequence, the mutant shifts based on the UNPROT1 model are in better agreement with experiment than referenced to the PROT1 or PROT1b model. Certainly, also these results should not be considered as final proof of a charged Glu181. However, previous arguments for a neutral Glu181 based on the S_1–S_2 energy gap measured in the two-photon absorption spectrum[62,128,129] are clearly disproven by our calculations. Again, the absolute energy of this state cannot be predicted with sufficient accuracy by state-of-the-art computational methods. The present calculations, however, show that the S_1–S_2 gap is hardly affected by the protonation state of Glu181, in agreement

with recent CASPT2 studies.[65,91]

All arguments in favor of a protonated Glu181 based on Raman and UV–vis data assume a pronounced effect of Glu181 mutations that should be detectable with theoretical methods. No such pronounced effects were found in the calculations neither for the unprotonated nor for the protonated case. Hence, the arguments of these works (Ref. 11 for Raman and Ref. 57 for UV–Vis) do not hold.

Further experimental evidence for an unprotonated Glu181 comes from the measured reactivity toward hydroxylamine.[57] Clearly, the calculation of this reaction is beyond current computational possibilities. However, the reaction requires hydroxylamine to enter the binding site, which requires a certain flexibility in the HBN around the PSB. The latter is most stable in case of a negatively charged Glu181, as inferred from dynamic data and hydrogen bond lengths. Models with protonated Glu181 show a similar stability of the HBN as the E181Q and E181A mutants. The measured reactivity, which is drastically increased in all Glu181 mutants except E181D, is more consistent with the UNPROT1 model.

In contrast to UV–vis and retinal stretch frequencies, NMR chemical shifts are sensitive to local fluctuations of the ESP and can in principle detect the presence of a charge at the position of residue 181. In lack of a reference like the E181Q mutant, the NMR measurements are difficult to interpret, because nearby polar groups or a displaced counterion may compensate or substitute the effect of the more distant charge monopole. The calculated ^{13}C shifts for 11-*cis* retinal in WT-Rh and the *trans* isomer in bR show significantly smaller errors (RMSD of 1.8 and 2.2 ppm, respectively) than previous theoretical works[88,89,137] and were able to identify systematic deviations between the WT-Rh models with protonated and unprotonated Glu181. These deviations were shown to be independent from details of the computational approach and fully account for the larger deviation in the PROT1 and PROT1b models (RMSD 3.1 ppm) compared to experiment. Still, the comparison of absolute shifts against experiment provides no satisfying evidence due to the remaining uncertainties of the QM/MM model (resolution of the X-ray structure, point charge model). More stringent are the results for the E181Q mutant. Relative to the UNPROT1 model, strong downfield shifts for the mutant at the C_5, C_7, and C_9 positions, and upfield shifts on C_{12} and C_{13} are obtained, which are absent in the PROT1 and PROT1b models. One of these pronounced shifts (C_{12}) has indeed been measured recently,[71] in good agreement with the UNPROT1 model. In comparison with the calculations, measurements for the remaining carbons would further improve the state of evidence.

Based on a comparison of several WT-Rh models, the predicted absolute values for all models and spectroscopic methods (except IR) are still compatible with both protonation states, however, small but systematic trends are found in the data in favor of an unprotonated Glu181. More specifically, the results obtained with the UNPROT1 model agree slightly but consistently better with experiment, their deviations in mutant shifts and NMR chemical shifts being smaller by a factor of 2 compared to those of the PROT1 and PROT1b models.

FTIR results have been interpreted as evidence for a unprotonated Glu181. The calculations clearly show, that a protonated Glu181 would absorb above $1700\,\mathrm{cm}^{-1}$, and therefore FTIR would find its signature in the difference spectra of dark state vs. later intermediates, in particular Meta I and Meta II, which have not been detected.[12] The possibility of a coupling of protonation states can be ruled out because the C=O vibrational frequncies of protonated Asp83 and Glu122 are nicely reproduced by the calculation and a band from a protonated Glu181 would be clearly distinguishable from Asp83 or Glu122 by position and shape, respectively.

Thus, the FTIR measurements of the carbonyl C=O stretch modes and the ^{13}C NMR chemical shift measurement of the C_{12} atom remain as the only published data that are indicative for the protonation state of Glu181. And both measurements are only consistent with the calculations assuming an unprotonated Glu181.

After the publication of these results and conclusions, the group of Birge et al. that have built the long-lasting image of a neutral binding pocket in the late 1980ies,[5,62] published a new work on the Glu181 protonation state.[138] They performed similar calculations with protonated and unprotonated Glu181 and the E181Q mutant in the batho state. They get a bathochromic shift due to the E181Q mutation of $-0.06\,\mathrm{eV}$ and this is only reproduced with their model with unprotonated Glu181 ($-0.007\,\mathrm{eV}$) and not with protonated Glu181 ($+0.16\,\mathrm{eV}$). With these results they conclude that Glu181 is unprotonated in the dark and batho state. They apologize in the last paragraph:

> *In closing, we note that a key conclusion of our previous two-photon study of rhodopsin was incorrect.*[62] *Failure to consider the possibility that one of the counterions could be hidden in a null point in the charge shift field led to the incorrect conclusion that the binding site is neutral. Conversely, it is interesting to note that the external point charge model of Honig and coworkers published in 1979 turns out to be surprisingly accurate.*[60] (from Ref. 138)

Chapter 5

Color Tuning Rh–HB

This chapter is about the third main topic of this work: the color tuning between rhodopsin and the human blue cone visual pigment. The similarity of rhodopsin and the cone pigments is already mentioned in the introduction. Du to the high simililarity of 41% between Rh and HB, conformational changes in the helices can be expected to be not very large. Thus, single and multiple mutations of some residues to their counterpart in HB are expected to give a sufficiently good structural model of HB to investigate the mechanism of color tuning. The main aspects are the interaction of the PSB with the complex counterion and electrostatic and steric interactions with polar groups in the binding pocket. Therefor, only mutations close to the chromophore are taken into account. In contrast to homology modeled structures of the full HB opsin, the result of this study is a binding pocket model. As a check for the quality of the model, shifts in excitation energies are calculated that should be similar to the experimental shift between Rh and HB of 0.5 eV.[18,20] The results of this chapter have in part been published in Ref. 139.

5.1 Model Building Methodology

5.1.1 Single Mutants

Starting with the bovine rhodopsin crystal structure (pdb code: 1U19)[52] and mutating some amino acids to the corresponding residues in the HB primary sequence, a model for the human blue (HB) visual pigment is built. Some mutations are very simple and a QM/MM structure optimization gives the only and thus best model structure. These kind of mutations are from small to small or from large to large amino acids. Especially mutations into a glycine or alanine are simple in respect to the residue conformation, but also with these mutations neighboring residues can change their orientation due to the mutation. For all mutations MD simulations were performed first to test the stability of the found

structure and second to find possible changes in the surrounding or even further away induced by small structural changes. Parallel to this, other structures and structures seen in the MD simulation were built by manually changing the orientation of the mutated or neighboring residues and again structure optimizations were performed.

The procedure of manually building the structures is favored against minimizing a snapshot from the MD because only local structural changes in the region of the mutation are modeled.

A manual evaluation of the structures can also give hints to structural changes in the neighboring residues. Especially missing or new hydrogen bonds are an indication for the orientation of polar groups. Mutations from large to small and/or apolar to polar residues can give the possibillity of a new water molecule being placed in the protein. The question if the choice of placing new water molecules is reasonable can not be answerd by energetic arguments but only by intuition.

All built structures for a specific mutation are evaluated with the total QM/MM energy and their stability in MD simulations. If the absorption maximum of a mutation has been investigated experimentally the shift in excitation energy relative to WT-Rh can be compared to the experimental shift to give another argument for or against a structure.

5.1.2 Multiple Mutants

To build the structures of multiple mutants the structures from single mutants or smaller multiple mutants were taken as templates. If a single mutant had more than one reasonably good structure, several structure were built for the multiple mutant. MD simulations were performed and again the structures were evaluated by energy, stability in the MD simulation and if an experimental shift exists by the calculated shift in excitation energy.

5.2 Choice of Mutations

The starting point of this work are the ten mutations investigated by Lin et al.[117] In this work from 1998 the selection of mutations was challenging. Without a crystal structure the only way to see if a residue is close to the chromophore was by inspection of the secondary structure model. With this small knowledge and the alignment of the primary structures, mutations that are thought to be close to retinal and to be important for color tuning were selected. Mainly mutation apolar–polar and vice versa were investigated. The twelve selected mutations are given in Table 26.

Table 26: Mutations selected by Lin et al.[117] and their experimental shift (relative to WT-Rh, eV), change in polarity, size and closest distance of sidechain atoms to retinal (Ret), PSB, or complex counterion (CCI) in WT-Rh (Å).

Mutant	Polarity	Size	$\Delta\Delta E_{S_1 \leftarrow S_0}$	d(Ret)	d(PSB)	d(CCI)
M86L	no→no	L→L	–		7.6	5.0
V87C	no→no	M→S	0.01		12.3	8.6
G90S	no→yes	XS→S	0.06		7.1	2.7
A117G	no→no	XS→XS	0.02	3.6	3.6	
E122L	yes→no	L→L	0.03	2.3[a]		
A124T	no→yes	XS→M	0.02	5.6[a]		
W265Y	yes→yes	XL→XL	0.08	3.5		
A292S	no→yes	XS→S	0.05	4.5	5.2	
A295S	no→yes	XS→S	0.03	4.3	6.5	
A299C	no→no	XS→S	0.01		9.9	8.9
P107R	no→yes	M→L	–	14.5		12.6
T108H	yes→yes	M→L	–	15.4		12.3

[a] Distance to β-ionone ring

The last two mutations were ignored in this work for four reasons: (1) They are very far away from the chromophore or complex counterion. (2) They have a very small shift in the work of Lin et al. (together 0.01 eV). (3) They are located at the edge of the transmembranal region and point in the direction where the membrane is in reality. (4) Arg107 and His108 are both titratable and can be protonated. To avoid these problems and because these two residues are expected to have a very small effect on the shift in excitation energy, they were neglected. The other ten mutants and some multiple mutants that were investigated by Lin et al. were built (see Table 27).

5.2.1 Search for More Amino Acids to be Mutated

The goal of this work is to build a binding pocket model of the blue cone visual pigment that can be used to interpret the mechanism of the color tuning from WT-Rh to HB. Therefore all mutations being close by and might have a structural influence on the chromophore have to be included. All residues being further away but having an impact on the excitation energy should also be included. To search for the most important mutations a perturbation analysis for all amino acids is performed. Therefore, in one calculation for each residue the partial charges of this residue were deleted to mimic the effect of a mutation into a glycine. Without structure relaxation the excitation energies were calculated with the fast OM2/MRCI method. The electrostatic effect of residue I in structure X defined by

Table 27: Mutations included in multiple mutants investigated experimentally[117] and built in this work.

	helix 2	helix 3	helix 6	helix 7
h2	M86L G90S			
h2′	M86L V87C G90S			
h3		A117G E122L A124T		
h7				A292S A295S A299C
h27	M86L G90S			A292S A295S A299C
h36		A117G E122L A124T	W265Y	
h237	M86L G90S	A117G E122L A124T		A292S A295S A299C
h267	M86L G90S		W265Y	A292S A295S A299C
h267+117	M86L G90S	A117G	W265Y	A292S A295S A299C
h267+122	M86L G90S	E122L	W265Y	A292S A295S A299C
h267+124	M86L G90S	A124T	W265Y	A292S A295S A299C
9fold	M86L G90S	A117G E122L A124T	W265Y	A292S A295S A299C
10fold	M86L V87C G90S	A117G E122L A124T	W265Y	A292S A295S A299C

$$ES^X(I) = \Delta E_{S_1 \leftarrow S_0}(X) - \Delta E_{S_1 \leftarrow S_0}(X \text{ w/o charges of } I) \quad (31)$$

Residues that have an electrostatic effect of $\left|ES^{\text{WT-Rh}}(I)\right| > 0.004\,\text{eV}$ and are different in the WT-Rh and HB primary sequence were first modeled as single mutants and then incorporated in a multiple mutant. The results of this perturbation analysis is given in Section 5.3.1.2. 14 residues were selected and added to the 10fold mutant to give the 24fold mutant (see Table 28).

Table 28: 24fold multiple mutant. 14 new mutations marked in **bold**.

helix 1	helix 2	helix 3	helix 4	loop E2	helix 5	helix 6	helix 7
Y43F	M86L V87C G90S **T92F** **T93P** **T94V** **T97A**	**N111A** A117G E122L A124T	**C167V**	**Y178F** **Y191W**	**Y206F** **M207L** **H211C**	W265Y	A292S A295S **T297S** **S298A** A299C

The search for other close by mutations was performed in an area of 5 Å around retinal and Lys296 sidechain. This region consists of 43 amino acids, 23 of them are different in WT-Rh and HB. 17 of these mutations are already included in the 24fold mutant and thus six new mutations were selected, modeled as single mutants and then added to the 24fold mutant to give the 30fold mutant.

Table 29: 30fold multiple mutant. Six new mutations marked in **bold**.

helix 1	helix 2	helix 3	helix 4	loop E2	helix 5	helix 6	helix 7
Y43F L47V **L47V**	M86L V87C G90S **F91V** T92F T93P T94V T97A	N111A A117G E122L A124T **L125G**	C167V	Y178F **I189P** Y191W	Y206F M207L H211C	W265Y **L266V**	A292S A295S T297S S298A A299C **V300I**

5.3 Structural Results of Model Building

5.3.1 Details of Single Mutants

Details of the single mutant structures are given to describe the orientation of the new residues and the changes of other residues due to the mutation. The location of the mutation is described with attributes right/left, above/below and in front/behind retinal. Therfore, retinal is placed with PSB on the right, β-ionone ring on the left and PSB N–H group pointing towards the viewer.

5.3.1.1 Ten Mutations investigated by Lin et al.

The first ten mutation are taken from the work of Lin et al.:[117] M86L, V87C, G90S, A117G, E122L, A124T, W265Y, A292S, A295S, A299C.

M86L mutation is located above PSB. This mutation is an exchange of a large residue to another large one and both residues are apolar. Thus, there is only a slight effect of the mutation on the surrounding residues (Phe91 is slightly displaced (< 0.5 Å))
Leu86 conformation: $\angle(\text{C–C}_\alpha\text{–C}_\beta\text{–C}_\gamma) = 54°$, $\angle(\text{C}_\alpha\text{–C}_\beta\text{–C}_\gamma\text{–H}_\gamma) = -70°$

V87C mutation is located left above PSB. Cys87 is slightly smaller than valine but there are no changes at the surrounding residues. Cys87 S–H points to the backbone oxygen atom of Asp83.
Cys87 conformation: $\angle(\text{C–C}_\alpha\text{–C}_\beta\text{–S}_\gamma) = 170°$, $\angle(\text{C}_\alpha\text{–C}_\beta\text{–S}_\gamma\text{–H}_\gamma) = 66°$

G90S mutation is located directly above Glu113, close to the position where Wat2b is located in WT-Rh. In this mutation a CH_2–OH group is newly induced and there are large changes in the complex counterion (Figure 30). Additionally, the CH_2 group of Phe116 is slightly displaced (1.2 Å) and hence the phenyl group of Phe116 is rotated but not displaced. Ser90 O–H points to the backbone oxygen atom of Leu112.
Ser90 conformation: $\angle(\text{C–C}_\alpha\text{–C}_\beta\text{–O}_\gamma) = 168°$, $\angle(\text{C}_\alpha\text{–C}_\beta\text{–O}_\gamma\text{–H}_\gamma) = -70°$

A117G mutation is located directly above the C_{12}–C_{13} single bond of retinal. The missing CH_3 group leaves a cavity for an extra water, that is added to the 10fold and other multiple mutants. There are no other changes at residues near by.

E122L mutation is located directly in front of the β-ionone ring, close to the C_4–C_5 single bond. Glu122 is protonated in WT-Rh and has a hydrogen bond to the backbone oxygen atom of His211. Leu122 is apolar but it pushes Wat6 slightly towards Trp126 that is rotated and has a hydrogen bond with Wat6 (Figure 31). Also His211 is slightly rotated.
Leu122 conformation: $\angle(\text{C–C}_\alpha\text{–C}_\beta\text{–C}_\gamma) = 150°$, $\angle(\text{C}_\alpha\text{–C}_\beta\text{–C}_\gamma\text{–H}_\gamma) = -165°$

A124T mutation is located far above retinal, ≈ 8 Å above the β-ionone ring. The newly induced CH_3 group does not perturb nearby residues but the O–H group

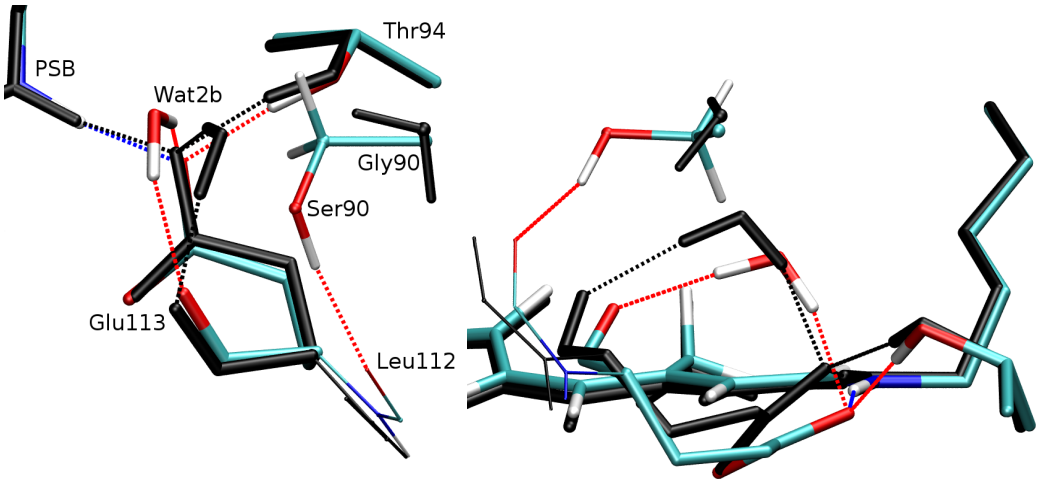


Figure 30: Orientation of Ser90 in the G90S single mutant in comparison to the WT-Rh structure (black) from two different viewpoints.

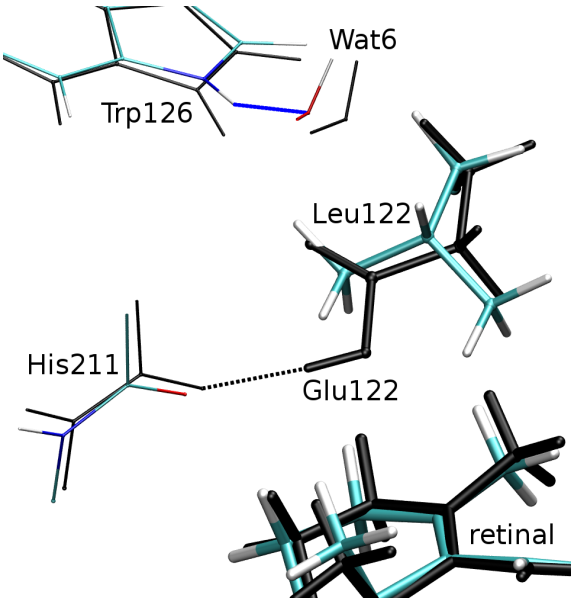


Figure 31: Orientation of Leu122 in the E122L single mutant in comparison to the WT-Rh structure (black).

points to the backbone oxygen atom of Gly120. Wat21 (hydrogen bonded to the backbone oxygen atom of Gly120 in WT-Rh) is rotated to have a hydrogen bond to Thr124 (Figure 32). Other changes are small (Leu79 and Wat15 are translocated by ≈ 0.5 Å).
Thr124 conformation: $\angle(\mathrm{C}\text{–}\mathrm{C}_\alpha\text{–}\mathrm{C}_\beta\text{–}\mathrm{O}_\gamma) = -176°$, $\angle(\mathrm{C}_\alpha\text{–}\mathrm{C}_\beta\text{–}\mathrm{O}_\gamma\text{–}\mathrm{H}_\gamma) = 73°$

W265Y mutation is located behind retinal, ≈ 5 Å away from C_8. Tyr265 is a bit smaller and the polar group (N–H in Trp265, and O–H in Tyr265) is moved to the right. The difference in size gives enough space for an extra water put into the structure. This Wat45 is placed between Tyr265 O–H and retinal (3.6 Å from C_{10}). It is hydrogen bonded to Tyr265 O–H and the backbone oxygen atom of Ala117 (Figure 33). The missing N–H group gives the formerly hydrogen bonded Wat14 the freedom to move away (1.7 Å) and Phe261 can move in this cavity (movement

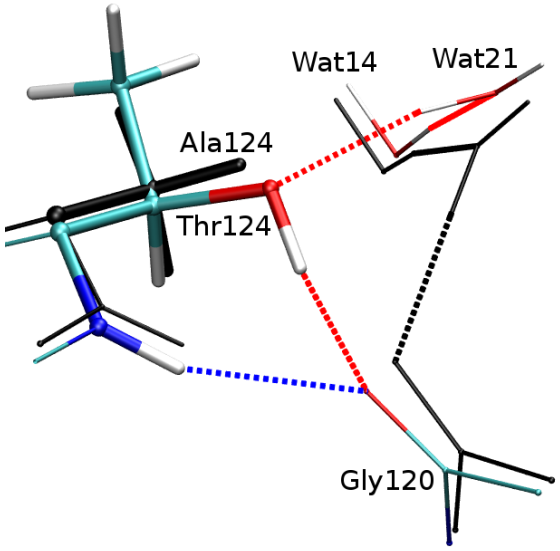


Figure 32: Orientation of Thr124 in the A124T single mutant in comparison to the WT-Rh structure (black).

of ≈ 1 Å). The new O–H group builds a hydrogen bond to Ser298 O–H that is strongly rotated to have itself a hydrogen bond to Wat21 instead of the backbone oxygen atom of Ala295. Wat21 itself is translocated by 1.1 Å.
Tyr265 conformation:
$\angle(\mathrm{C}\text{–}\mathrm{C}_\alpha\text{–}\mathrm{C}_\beta\text{–}\mathrm{C}_\gamma) = 149°$, $\angle(\mathrm{C}_\alpha\text{–}\mathrm{C}_\beta\text{–}\mathrm{C}_\gamma\text{–}\mathrm{C}_{\delta 1}) = 94°$, $\angle(\mathrm{C}_{\epsilon 1}\text{–}\mathrm{C}_\zeta\text{–}\mathrm{O}_\eta\text{–}\mathrm{H}_\eta) = -23°$

Tyr265 and Ala124 are only 5 Å away from each other and both mutations (A124T, W265Y) have a structural effect on Wat21. Furthermore, both mutations together are largely different to both single mutations (see Section 5.3.2.1).

A292S mutation is located directly below retinal, 3.3 Å from C_{15}. Ser292 O–H has a hydrogen bond to Wat2a and it is hydrogen bonded by Ser186 (hydrogen bonded to Wat2a in WT-Rh) (Figure 34).
Ser292 conformation: $\angle(\mathrm{C}\text{–}\mathrm{C}_\alpha\text{–}\mathrm{C}_\beta\text{–}\mathrm{O}_\gamma) = 53°$, $\angle(\mathrm{C}_\alpha\text{–}\mathrm{C}_\beta\text{–}\mathrm{O}_\gamma\text{–}\mathrm{H}_\gamma) = 102°$

A295S mutation is located behind retinal, 3.3 Å from the methyl group at C_{13}. There are three different orientations for the Ser295 O–H group (a,b, and c). For all of them there are no major changes at other residues.

In the *A295S_a* structure the Ser295 O–H has a hydrogen bond to the backbone oxygen atom of Ala292.
Ser295 conformation: $\angle(\mathrm{C}\text{–}\mathrm{C}_\alpha\text{–}\mathrm{C}_\beta\text{–}\mathrm{O}_\gamma) = -54°$, $\angle(\mathrm{C}_\alpha\text{–}\mathrm{C}_\beta\text{–}\mathrm{O}_\gamma\text{–}\mathrm{H}_\gamma) = -50°$

In the *A295S_b* structure the Ser295 O–H has a hydrogen bond to the backbone oxygen atom of Pro291.
Ser295 conformation: $\angle(\mathrm{C}\text{–}\mathrm{C}_\alpha\text{–}\mathrm{C}_\beta\text{–}\mathrm{O}_\gamma) = -76°$, $\angle(\mathrm{C}_\alpha\text{–}\mathrm{C}_\beta\text{–}\mathrm{O}_\gamma\text{–}\mathrm{H}_\gamma) = -77°$

In the *A295S_c* structure the Ser295 O–H has a hydrogen bond to the backbone oxygen atom of Pro291 and Wat8 is rotated to have a hydrogen bond to Ser295 O–H.
Ser295 conformation: $\angle(\mathrm{C}\text{–}\mathrm{C}_\alpha\text{–}\mathrm{C}_\beta\text{–}\mathrm{O}_\gamma) = 177°$, $\angle(\mathrm{C}_\alpha\text{–}\mathrm{C}_\beta\text{–}\mathrm{O}_\gamma\text{–}\mathrm{H}_\gamma) = 76°$

Table 30 shows the arguments for choosing A295S_a as the structure for all mul-

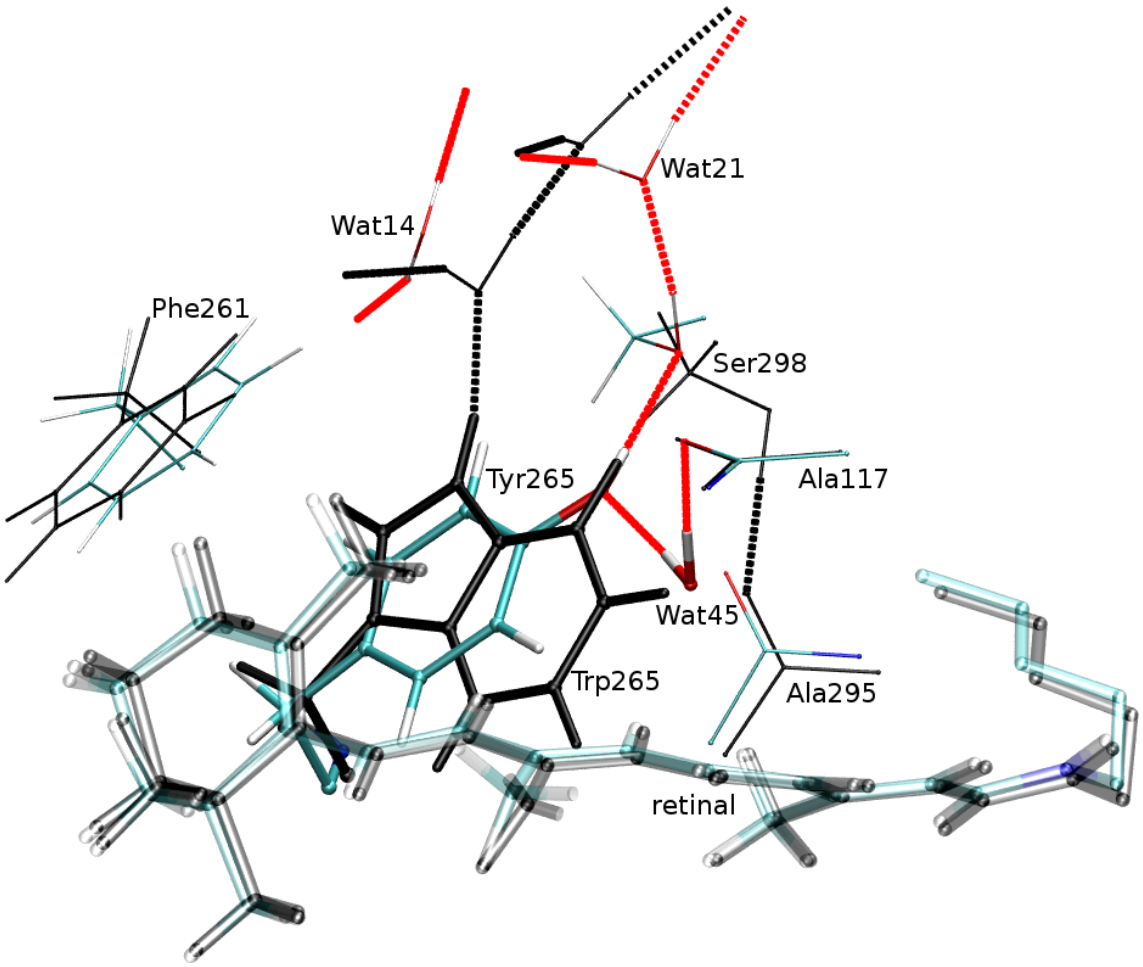


Figure 33: Orientation of Tyr265 in the W265Y single mutant in comparison to the WT-Rh structure (black).

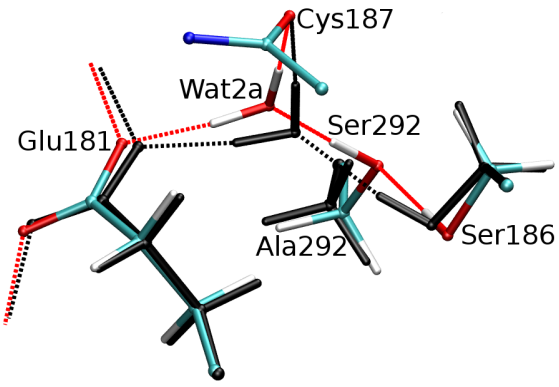


Figure 34: Orientation of Ser292 in the A292S single mutant in comparison to the WT-Rh structure (black).

tiple mutants: A295S_b is too high in energy and A295S_c has a wrong shift in excitation energy (-0.06, exp.: $+0.03\,\text{eV}$). The introduction of a new water molecule close to Ser295 has been tried, but in MD simulation no stable structure was found. Thus, A295S_a was chosen for the single mutant and as template for multiple mutants.

Table 30: Energetical difference ΔE (relative to A295S_a, DFTB/CHARMM, kcal/mol) and shift in excitation energy $\Delta\Delta E_{S_1 \leftarrow S_0}$ (relative to WT-Rh, SORCI, eV (exp.: 0.03 eV)) of the three A295S mutant structures.

	A295S_a	A295S_b	A295S_c
ΔE	—	0.9	-0.3
$\Delta\Delta E_{S_1 \leftarrow S_0}$	-0.01	0.05	-0.06

A299C mutation is located far above the PSB ($\approx 10\,\text{Å}$). The Cys299 S–H points towards the backbone oxygen atom of Asp83, but it is too far away to have a hydrogen bond ($d(\text{S–O}) = 3.6\,\text{Å}$). Val87 is slightly moved by 0.8 Å due to the large cysteine.
Cys299 conformation: $\angle(\text{C–C}_\alpha\text{–C}_\beta\text{–S}_\gamma) = 64°$, $\angle(\text{C}_\alpha\text{–C}_\beta\text{–S}_\gamma\text{–H}_\gamma) = 76°$

The V87C and A299C mutations are very close to each other ($d(\text{S–S}) = 2.6\,\text{Å}$ in the single mutants) and both point in the same direction in the single mutants. In multiple mutants having both mutations the orientation will be different (see Section 5.3.2.1).

5.3.1.2 14 Mutations Found by Perturbation Analysis

Perturbation analysis was performed for nearly all of the 348 amino acids in WT-Rh. The first and last residue were not tested, because they are far away (26 and 47 Å for Met1 and Ala348, respectively) and they are both charged due to the terminating capping group. Of course glycine residues were not tested because the perturbation works like a mutation into a glycine and there would be no change. Additionally, Lys296, Glu113, and Thr94 were not tested because they are part of the QM4 region. Thr94 is important because it is mutated into a valine in HB. Thus, the electrostatic effect of Thr94 is tested in the QM2 region.

The results of the perturbation analysis are given in Table 31.

The 14 mutations with an electrostatic effect of $\left|ES^{\text{WT-Rh}}(I)\right| > 0.004\,\text{eV}$ are modeled as single mutants and their structures are described in the following.

Y43F mutation is located right of PSB ($\approx 5\,\text{Å}$ away from Lys296 side chain) at the outside of the protein. Due to the missing Tyr43 O–H group and its hydrogen bond to the backbone oxygen atom of Phe293, Phe43 is slightly rotated in comparison to Tyr43.
Phe43 conformation: $\angle(\text{C–C}_\alpha\text{–C}_\beta\text{–C}_\gamma) = 71°$, $\angle(\text{C}_\alpha\text{–C}_\beta\text{–C}_\gamma\text{–C}_{\delta 1}) = 80°$

Table 31: Results of the perturbation analysis ($ES^{\text{WT-Rh}}(I)$, OM2/MRCI, eV). Mutations already included in the 10fold mutant are given in **bold**.

WT-Rh	HB	$ES^{\text{WT-Rh}}(I)$	WT-Rh	HB	$ES^{\text{WT-Rh}}(I)$
Ala117	**Gly**	−0.003	Trp126	–	0.004
Cys185	–[a]	−0.004	Tyr206	Phe	0.004
Tyr301	–	−0.005	**Glu122**	**Leu**	0.004
Gln184	–	−0.005	Thr92	Phe	0.004
His100[b]	Asn	−0.005	Ser127	–	0.005
His152	–	−0.005	Thr289	–	0.005
Met207	Leu	−0.005	Phe261	–	0.005
Asn55	–	−0.006	Ser98	–	0.007
Thr93	Pro	−0.006	Met44	–	0.009
Thr97	Ala	−0.007	Ser298	Ala	0.009
Asn111	Ala	−0.008	His211	Cys	0.014
Thr178	Phe	−0.008	Phe293	–	0.019
Thr297	Ser	−0.011	Glu181	–	0.064
Cys167	Val	−0.012	Mutations from 10fold mutant:		
Tyr43	Phe	−0.018	**Met86**	**Leu**	0.001
Thr94[c]	Val	−0.020	**Val87**	**Cys**	0.000
Tyr268	–	−0.022	**Gly90**	**Ser**	—
Trp265	**Tyr**	−0.024	**Ala124**	**Thr**	0.000
Tyr191	Trp	−0.025	**Ala292**	**Ser**	0.000
Ser186	–	−0.031	**Ala295**	**Ser**	0.002
Thr118	–	−0.043	**Ala299**	**Cys**	0.000

[a] Conserved residue in WT-Rh and HB.

[b] His100 is located more than 16 Å away from the PSB at the outer surface of WT-Rh and thus this mutation is not modeled.

[c] Thr94 is part of the QM4 region and is tested with the QM2 region.

T92F mutation is located right, in front of the complex counterion, ≈ 10 Å away from Glu113. Due to the missing Thr92 O–H group, Wat19 is slightly moved (1.2 Å) and Thr93 is moved to have a hydrogen bond to Wat19 (Figure 35).
Phe92 conformation: $\angle(\text{C–C}_\alpha\text{–C}_\beta\text{–C}_\gamma) = 66°$, $\angle(\text{C}_\alpha\text{–C}_\beta\text{–C}_\gamma\text{–C}_{\delta 1}) = 74°$

T93P mutation is located right, in front of the complex counterion, ≈ 6 Å away from Glu113. Due to the missing Thr93 O–H group, Wat19 is slightly moved (1.5 Å) and due to the missing N–H group, the backbone oxygen atom of Gly90 is moved apart (0.7 Å) (Figure 36).
Pro93 conformation: $\angle(\text{C–C}_\alpha\text{–C}_\beta\text{–C}_\gamma) = -99°$, $\angle(\text{C}_\alpha\text{–C}_\beta\text{–C}_\gamma\text{–C}_\delta) = -34°$

T92F and T93P mutations are very close to each other (≈ 3 Å) and both have an influence on Wat19. Multiple mutants having both mutations will differ in this region from the single mutants.

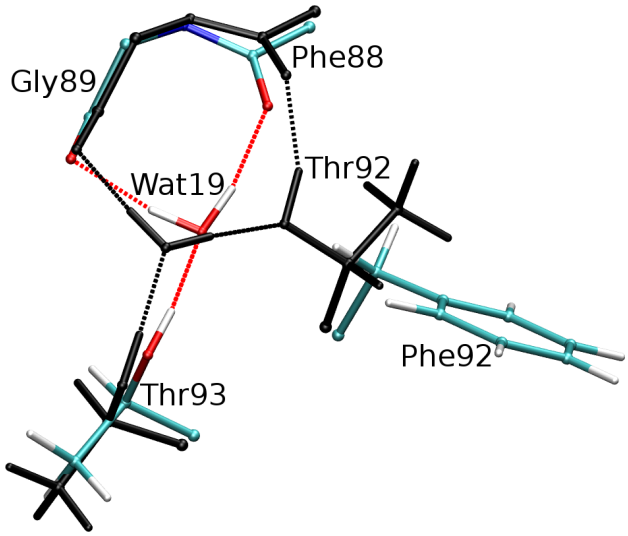


Figure 35: Orientation of Phe92 in the T92F single mutant in comparison to the WT-Rh structure (black). Retinal is located at the bottom left, ≈ 11 Å away from residue 92.

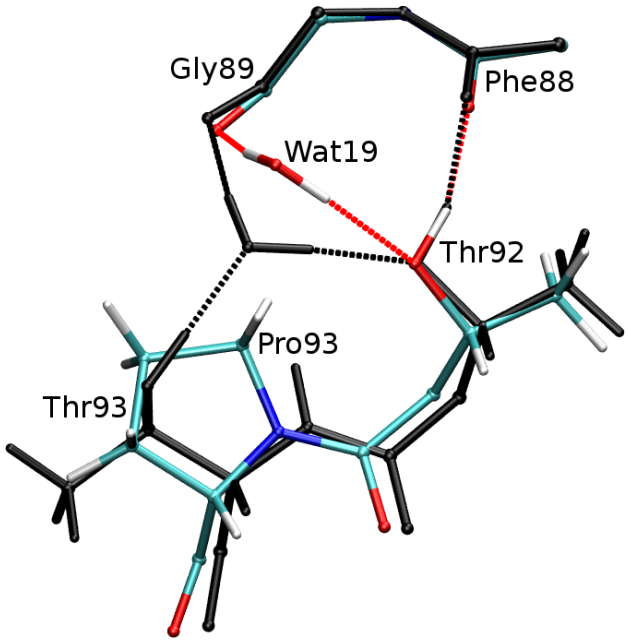


Figure 36: Orientation of Pro93 in the T93P single mutant in comparison to the WT-Rh structure (black). Retinal is located at the bottom left, ≈ 11 Å away from residue 93.

T94V mutation is located directly at the complex counterion. Thr94 is hydrogen bonded to Glu113 and is part of the QM4 region. In the T94V single mutant and all multiple mutants containing this mutation, the valine side chain is part of the QM4 region. Threonine and valine are similar in size but due to the missing Thr94 O–H group and its hydrogen bond to Glu113 the Val94 side chain is differently oriented and thus the backbone is disturbed. The C_α and nitrogen atom of Val94 are moved down by 0.5 and 0.4 Å, respectively. The influence of this mutation on other residues is small and mainly results from the perturbation of the backbone.
Val94 conformation: $\angle(\text{C–C}_\alpha\text{–C}_\beta\text{–H}_\beta) = 163°$

Other structures for the T94V mutation were built by rotating around the C_α–C_β bond, but give higher energies mainly due to unfavorable orientation of the methyl groups to Glu113.

T97A mutation is located left, in front of Glu113, ≈ 7 Å away. Although Ala97 is much smaller than threonine, there are no changes at other residues except small perturbations in the backbone of helix 2 due to the missing hydrogen bond from Thr97 O–H group to the backbone oxygen atom of Thr93.

N111A mutation is located at the outside of the protein in front of retinal, 11 Å from C_{12}. Ala111 is much smaller than asparagine but the changes of other residues are limited. Only Wat38 (hydrogen bonded by the N_δ–$H_{\delta 1}$ group of Asn111 in WT-Rh) is displaced by 1 Å and due to this movement Wat23 is slightly rotated. Wat11 (hydrogen bonded to Asn111 O_δ in WT-Rh) is slightly rotated to have a hydrogen bond to the backbone oxygen atom of Pro171.

This mutation from a large to a very small residue is a good example for addition of an extra water molecule into the free cavity produced by the mutation. In the mutant structure N111A_b, an extra water is placed on the position of the NH_2 group of Asn111. In the optimized structure this water molecule is slightly moved to the outside of the protein (1.4 Å) and prevents Wat38 to move apart with a hydrogen bond. But these changes only have a very limited effect on the excitation energy ($\Delta\Delta E_{S_1 \leftarrow S_0} = 0.03$ eV and 0.02 eV for N111A and N111A_b, respectively) and due to the fact that Ala111 and Wat38 are located at the very edge of the protein, this extra water is not added to multiple mutants and the N111A mutant structure is used for the 24fold and higher multiple mutants.

C167V mutation is located in front of the β-ionone ring, 7 Å from C_7. Val167 is larger than cystein and the larger sterical interactions move Met207 (0.5 Å) and rotates Wat6 and His211. Met207 and His211 are also mutated in the 24fold mutant.
Val167 conformation: $\angle(\text{C–C}_\alpha\text{–C}_\beta\text{–H}_\beta) = -74^\circ$

Y178F mutation is located in front of retinal, ≈ 6 Å from C_{11}. It has no influence on other residues.
Phe178 conformation is the same as Tyr178 in WT-Rh: $\angle(\text{C–C}_\alpha\text{–C}_\beta\text{–C}_\gamma) = 166^\circ$, $\angle(\text{C}_\alpha\text{–C}_\beta\text{–C}_\gamma\text{–C}_{\delta 1}) = -69^\circ$

Y191W mutation is located below retinal, ≈ 5 Å from the methyl group at C_9. The plane of the indole group has the same orientation as the phenyl group of Tyr191 in WT-Rh but the conformation at the C_α–C_β bond is clearly changed. Due to this change, the C_α atom is moved by 0.5 Å and due to the missing hydrogen bond of the Tyr191 O–H group to Tyr268 in WT-Rh, Tyr268 has more flexibility and moves 0.7 Å apart. The N_ϵ–H_ϵ group of Trp191 has a hydrogen bond to the negatively charged Glu181 $O_{\epsilon 1}$ atom (Figure 37). Ile189 and Met207 are slightly rotated.
Trp191 conformation: $\angle(\text{C–C}_\alpha\text{–C}_\beta\text{–C}_\gamma) = -90^\circ$, $\angle(\text{C}_\alpha\text{–C}_\beta\text{–C}_\gamma\text{–C}_{\delta 1}) = 64^\circ$

Y206F mutation is located left, in front of the β-ionone ring, ≈ 11 Å away from C_6. Due to the missing hydrogen bond from the O–H group of Tyr206, His211 slightly moves apart (0.4 Å).
Phe206 conformation: $\angle(\text{C–C}_\alpha\text{–C}_\beta\text{–C}_\gamma) = 39^\circ$, $\angle(\text{C}_\alpha\text{–C}_\beta\text{–C}_\gamma\text{–C}_{\delta 1}) = 74^\circ$

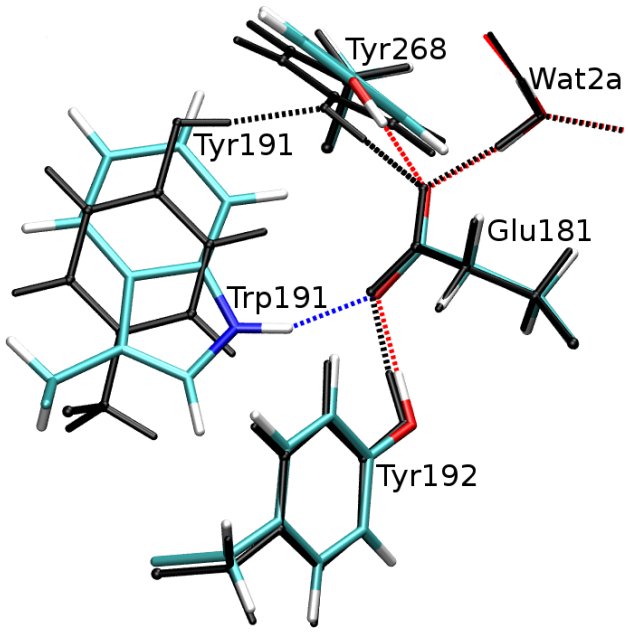


Figure 37: Orientation of Trp191 in the Y191W single mutant in comparison to the WT-Rh structure (black).

M207L mutation is located below, in front of the β-ionone ring, ≈ 5 Å from C_7. Both amino acids are of similar size and thus only the S–H group of Cys167 is moved apart by 0.4 Å.
Phe206 conformation: $\angle(C\text{–}C_\alpha\text{–}C_\beta\text{–}C_\gamma) = 56°$, $\angle(C_\alpha\text{–}C_\beta\text{–}C_\gamma\text{–}H_\gamma) = -53°$

H211C mutation is located left, in front of the β-ionone ring, ≈ 8 Å from C_6. Cys211 is much smaller than histidine, and thus Wat6 is moved by 2.2 Å to be hydrogen bonded by the S–H group of Cys211. Glu122 has more flexibility and gets a hydrogen bond from the N_ϵ–H_ϵ group of Trp126. Tyr206 has a hydrogen bond to the $N_{\epsilon 2}$ atom of His211 in WT-Rh and is rotated to have a hydrogen bond to the backbone oxygen atom of Ala166 in the H211C mutant structure (Figure 38).
Cys211 conformation: $\angle(C\text{–}C_\alpha\text{–}C_\beta\text{–}S_\gamma) = -173°$, $\angle(C_\alpha\text{–}C_\beta\text{–}S_\gamma\text{–}H_\gamma) = -64°$

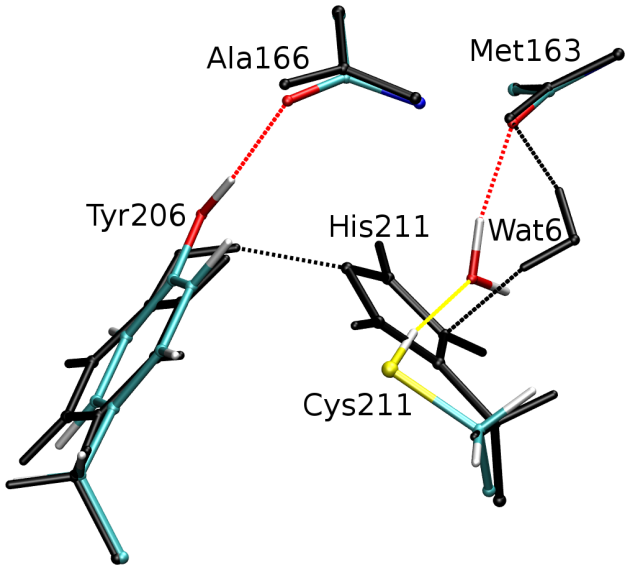


Figure 38: Orientation of Cys211 in the H211C single mutant in comparison to the WT-Rh structure (black).

T297S mutation is located behind, ≈ 10 Å above PSB. The orientation of the O–H group is the same as of Thr297 in WT-Rh.
Ser297 conformation: $\angle(\text{C–C}_\alpha\text{–C}_\beta\text{–O}_\gamma) = -64°$, $\angle(\text{C}_\alpha\text{–C}_\beta\text{–O}_\gamma\text{–H}_\gamma) = -82°$

S298A mutation is located ≈ 8 Å above PSB. There is no influence on other residues.

5.3.1.3 Six Mutations Found in 5 Å Around Retinal

Analysis of the sequence of WT-Rh and HB give all differences. Mutations closer than 5 Å to retinal or Lys296 side chain are listed in Table 32. Many of these mutations are already included in the 10fold or 24fold mutant. The remaining six mutations are included in the 30fold mutant and are described in the following.

Table 32: Mutations close to retinal and Lys296 side chain and their inclusion in multiple mutants.

Mutation	included in	Mutation	included in	Mutation	included in
Y43F	24, 30	L125G	30	L266V	30
L47V	30	C167V	24, 30	A292S	10, 24, 30
M86L	10, 24, 30	Y178F	24, 30	A295S	10, 24, 30
G90S	10, 24, 30	I189P	30	T297S	24, 30
F91V	30	Y191W	24, 30	S298A	24, 30
T94V	24, 30	M207L	24, 30	A299C	10, 24, 30
A117G	10, 24, 30	H211C	24, 30	V300I	30
E122L	10, 24, 30	W265Y	10, 24, 30		

An increase of the region to 6 Å gives ten additional mutations (L40Q, L119V, G120A, I123V, A124T, A168S, F203Y, G270A, F273M, and M288V). One of those, A124T, is already included in the 10fold mutant. Six mutations included in the 24fold mutant are even further away than 6 Å (V87C, T92F, T93P, T97A, N111A, and Y206F). A 39fold mutant is thought to give no new insight in the binding pocket around retinal and is thus not modeled.

L47V mutation is located ≈ 6 Å right of the C_α atom of Lys296. Valine is smaller than leucine but the small displacement of Phe91 (0.5 Å) results from the new, second methyl group at the C_β atom of Val47.
Val47 conformation: $\angle(\text{C–C}_\alpha\text{–C}_\beta\text{–H}_\beta) = -75°$

F91V mutation is located right, ≈ 6 Å above Glu113. Ala299 has more flexibility and moves down by 0.8 Å. Due to this movement, the backbone of Ala299 is disturbed resulting in a small rotation of Asp83 and a movement of Wat15 (1 Å). Val91 and Wat15 are ≈ 12 Å away from each other. Thus, this is a very good example for possible long-range effects of a single mutation. Due to the interaction with the second methyl group, Ile48 is displaced by 0.9 Å.
Val91 conformation: $\angle(\text{C–C}_\alpha\text{–C}_\beta\text{–H}_\beta) = -71°$

L125G mutation is located left, above the β-ionone ring, ≈ 5 Å from C_4. Gly125 is much smaller than leucine but the effect of this single mutation on other residues is very small. Only Phe261 is slightly rotated (C_ζ is moved 0.6 Å).

I189P mutation is located in front of retinal, ≈ 3 Å away from the methyl group at C_9. Pro189 is smaller than isoleucine but much more rigid and although both are apolar, there can be an electrostatic effect because the polar N–H group in the backbone of Ile189 is missing in Pro189. The mutation has an effect on some neighboring residues (Pro171, Ser176, Tyr191, Phe203, and Val204) and on the Pro189 backbone ($\phi = -71°$ ($-78°$ in WT-Rh), $\psi = 163°$ (150° in WT-Rh)). The methyl group at C_9 of retinal is moved apart by 0.3 Å and slightly rotated to have a distance $d(H_{\delta 2}–H_{19B})$ of 2.2 Å (without movement and rotation it would be only 1.9 Å).
Pro189 conformation: $\angle(C–C_\alpha–C_\beta–C_\gamma) = -91°$, $\angle(C_\alpha–C_\beta–C_\gamma–C_\delta) = -35°$

In the 30fold mutant L125G and I189P mutations together will be very important structurally (Section 5.3.2.4 and 5.4.2.4).

L266V mutation is located behind the β-ionone ring, ≈ 7 Å from C_2, at the outside of the protein. Phe212 is slightly rotated due to the missing sterical interaction with one of the Leu266 methyl groups.
Val266 conformation: $\angle(C–C_\alpha–C_\beta–H_\beta) = -73°$

V300I mutation is located right, above Lys296 side chain, ≈ 8 Å from C_α, at the outside of the protein. Thus, the effects on neighboring residues is small. Isoleucine is larger by one methyl group than valine, and thus Ile300 pushes the sidechain of Leu50 0.5 Å away.
Ile300 orientation: $\angle(C–C_\alpha–C_\beta–H_\beta) = 53°$, $\angle(C_\alpha–C_\beta–C_{\gamma 1}–C_\delta) = -172°$

5.3.2 Details of Multiple Mutants

Structural changes in the multiple mutants in comparison to the single mutants are described. Orientations that are the same in the single mutant or in the smaller multiple mutants are not listed.

5.3.2.1 Multiple Mutants by Lin et al.

Lin et al.[117] have built several multiple mutants for every helix (h2, h2′, h3, and h7) and combination of these resulting in a 9fold and 10fold mutant.

The **h2 multiple mutant** consists of M86L and G90S mutations. Additionally, the **h2′ multiple mutant** has the V87C mutation. In both multiple mutants there are no cooperative effects of the mutations. All residues are oriented as in WT-Rh or in the single mutants if changed.

Also for the **h3 multiple mutant** there are no cooperative effects. The three mutations A117G, E122L, and A124T are far away from each other and all residues are oriented as in WT-Rh or in the single mutants if changed.

The same is true for the **h7 multiple mutant** consisting of the A292S, A295S, and A299C mutations.

The last multiple mutant without cooperative effects is the **h27 multiple mutant**. It is a combination of h2 and h7. All five mutations (M86L, G90S, A292S, A295S, A299C) are sufficiently seperated and only have local effects on other residues.

In the **h36 multiple mutant** some residues are differently oriented than in the h3 and W265Y (=h6) mutants. Tyr265 and Ser298 are in the same orientation as in the W265Y single mutant. The newly induced Wat45 is slightly displaced due to the change in the Gly117 backbone (it is hydrogen bonded to Gly117 backbone oxygen). The water network around Asp83 and Thr124 is changed due to the rotation of Ser298 and the missing hydrogen bond from Trp265. Wat14 is strongly moved to the middle of the water network (3.8 Å) and Wat21 is displaced by 1.8 Å to a similar position as in the W265Y mutant structure. The difference is shown in Figure 39.

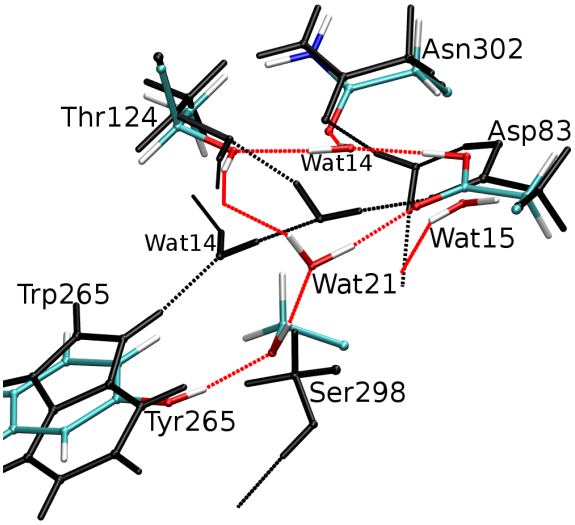


Figure 39: Water network around Asp83, Thr124, and Ser298 in the h36 mutant in comparison to the h3 mutant (black).

In the **h267 multiple mutant** a second water molecule is added. It is placed in the cavity built by the M86L mutation and it is hydrogen bonded to Wat45 (introduced with the W265Y mutation) and Wat2b (moved due to the G90S mutation). Wat2b is moved further towards Wat46 and thus Ser90 can rotate and build a hydrogen bond to the backbone oxygen atom of Glu113 instead of Leu112 (Figure 40).

Three multiple mutants more were built composed of the h267 multiple mutant and the three mutations in helix 3.

In the **h267+117 multiple mutant** Wat2b can move to the left due to the A117G mutation, and Ser90 rotates and points to Wat2b instead of the backbone oxygen atom of Glu113 as in the h267 multiple mutant (Figure 41).

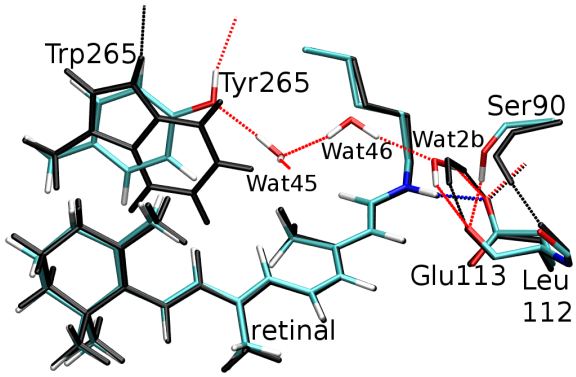


Figure 40: Change in the complex counterion and the HBN to Tyr265 in the h267 mutant in comparison to the h2 mutant (black).

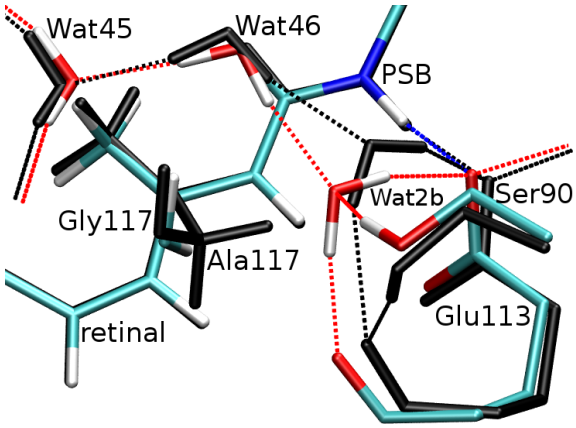


Figure 41: Change in the complex counterion in the h267+117 mutant in comparison to the h267 mutant (black).

In the **h267+122 multiple mutant** the E122L mutation induces the same changes as the single mutation but additionally the full retinal is moved to the direction of the β-ionone ring (0.3 Å). Additionally, Wat2b rotates and is hydrogen bonded to Ser90 instead of Glu113 backbone oxygen. But this is the only structure where this bonding situation is found.

In the **h267+124 multiple mutant** the changes introduced by the A124T mutation are the same as in the h36 multiple mutant (see Figure 39).

The **9fold mutant** can also be called h2367 multiple mutant and is simply an overlay of the h267+117 multiple mutant (in the complex counterion region), the E122L single mutant (in the β-ionone region) and the h36 multiple mutant (in the region of the water network around Thr124, Ser298, and Asp83).

In the **10fold mutant** (also called h2'367) the added mutation is V87C. Cys87 is oriented as in the single mutant (points to Asp83 backbone oxygen) and Cys299 (pointing also to Asp83 backbone oxygen in the 9fold mutant) is rotated by 90° ($\angle(C_\alpha\text{–}C_\beta\text{–}S_\gamma\text{–}H_\gamma) = 170°$).

The distance between the two sulfur atoms of Cys299 and Cys87 is 3.4 Å. It seems to be possible to build a disulfide bridge between these two residues. Nearly no difference in the shift in excitation energy (−0.01 eV) is observable. A disulfide bridge would perturb helices 2 and 7 due to decreased flexibility. Modelling this perturbation would need very long MD simulations. Additionally changes in helices would also perturb the outside of the protein which is not perfectly

modeled in the setup because of missing membrane and bulk water. Therefore, all higher multiple mutants are built without a disulfide bridge.

5.3.2.2 24fold Mutant

The 14 mutations found by perturbation analysis are added to the ten mutations of the 10fold mutant. Some mutated residues are very close together and thus the structure is different to a simple overlay of the single mutants. These cooperative effects are described in the following.

In the region in front of the β-ionone ring the E122L, C167V, and H211C mutations are very close and their orientation is different to the single mutations. Also the position of Wat6 and the orientation of Trp126 and Leu122 is changed (see Figure 42). Close to this region the M207L mutation is located. The orientation of Leu207 is not changed but Trp191 has more space to slightly rotate to the left.

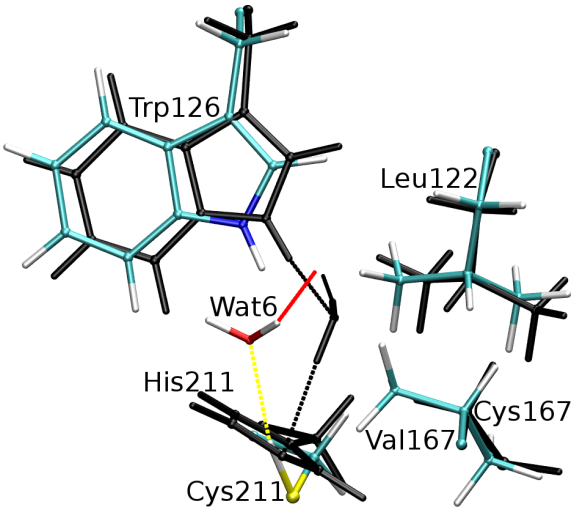


Figure 42: Changed orientation of Trp126, Leu122 and Wat6 in the 24fold mutant in comparison to the 10fold mutant (black).

In the region of the water network around Thr124, Ser298, and Asp83 the S298A mutation changes in the orientation of Wat21 and Tyr265 due to the missing hydrogen bond to these residues (Figure 43). The orientation being most stable in MD simulations contains Tyr265 O–H pointing in the direction of Wat21 but the distance is too far away for a hydrogen bond (d(Tyr265 O_η–Wat21 O) = 3.1 Å).

There are seven mutations in helix 2: M86L, V87C, G90S, T92F, T93P, T94V, and T97A. The last three mutations strongly perturb the backbone of helix 2. Due to this, the orientation of Phe92 is slightly changed in the 24fold mutant in comparison to the T92F single mutant (Figure 44).

Both cysteines, Cys87 and Cys299, change their orientation. The S–H group of Cys87 is pointing to the backbone oxygen atom of Asp83 as in the single mutant and the S–H group of Cys299 is rotated ($\angle$(C–C_α–C_β–S_γ) = 64°). This rotation can have various reasons (e.g. changes in backbone from helical movements, indirect sterical effects of other mutations). Both residues are $\approx$ 11 Å away from

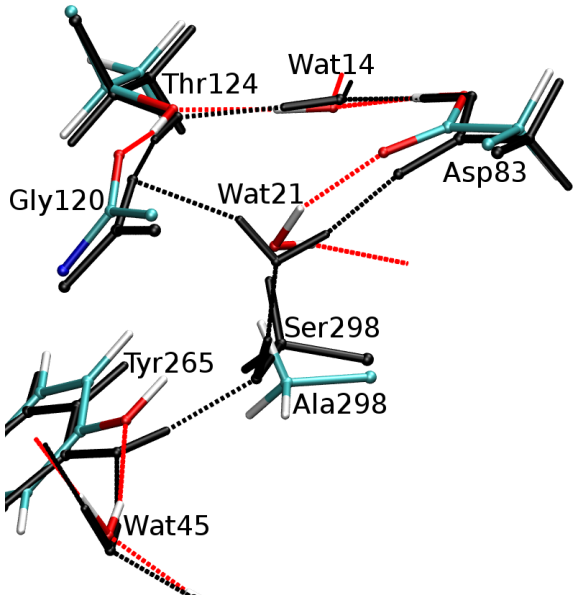


Figure 43: Changed orientation of Tyr265 and Wat21 in the 24fold mutant in comparison to the 10fold mutant (black).

PSB and thus a changed orientation does not play a role in color tuning electrostatically.

The 24fold mutant is stable in MD simulations for the first 200 ps. But after 200 ps the HBN between Tyr265 and Ser90 is disrupted and afterwards no other stable HBN was found in the MD simulation. Another structure for the 24fold mutant was built by adding a third water in the free space left by the S298A mutation. The HBN is completely changed and is very flexible. A slightly stable structure (24fold_b) is chosen but was discarded because of a very small shift in excitation energy (see Section 5.4.2.3).

5.3.2.3 30fold Mutant

The six new mutations are L47V, F91V, L125G, I189P, L266V, and V300I. Again some cooperative effects can be observed.

Leu122, Val167 and Cys211 are slightly moved due to the L125G mutation but their orientation is the same as in the 24fold mutant.

Ile300, Val47, and Val91 are very close to each other and together they influence the position of Cys299, Phe43, and Met44. Apart from the rotation of Phe43 the changes are very small and the conformation of these residues is the same as in the 24fold mutant.

The largest difference can be seen at retinal. It is moved in the direction of the β-ionone ring (to the left in Figure 45a) which has more space because of the missing sidechain of Leu125. Due to this movement, the sidechains of Phe212 and Val266 are slightly displaced but have the same orientation as in the 24fold mutant or the L266V single mutant. The origin of this large displacement of retinal is investigated in detail in the following section.

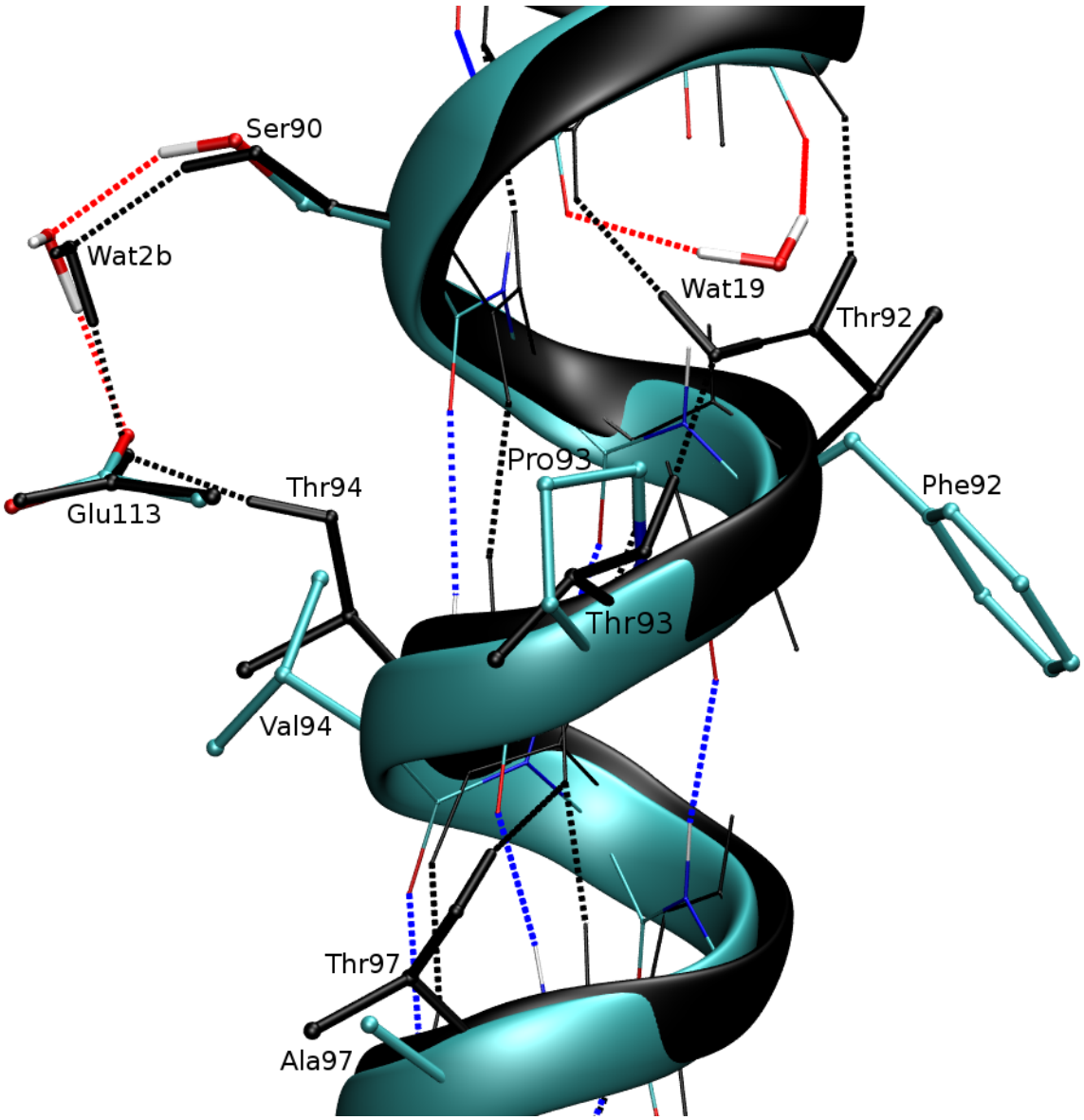


Figure 44: Changes in helix2 due to T93P, T94V, and T97A mutations in the 24fold mutant in comparison to the 10fold mutant (black).

5.3.2.4 Displacement of Retinal in the 30fold Mutant Due to L125G and I189P Mutations

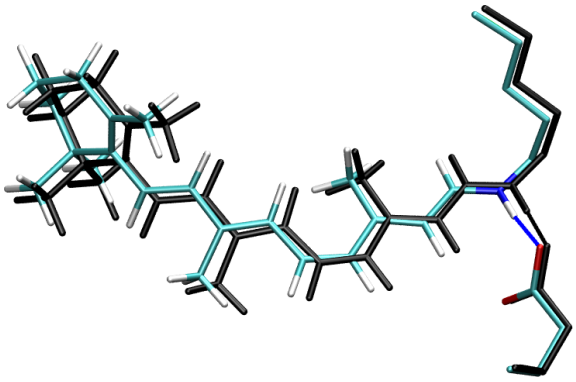
Three more structures were built, 28fold (≙30fold w/o L125G, I189P), 28fold+125 (≙30fold w/o I189P) and 28fold+189 (≙30fold w/o L125G), to investigate the origin of the large retinal displacement between the 24fold and 30fold mutant.

There is only a small difference between the 24fold and 28fold mutants in the β-ionone ring (due to the L266V mutation), and nearly no displacement at C_{19} (methyl group at C_9) and N_{16} (PSB) (Figure 45b and Table 33). The large displace-

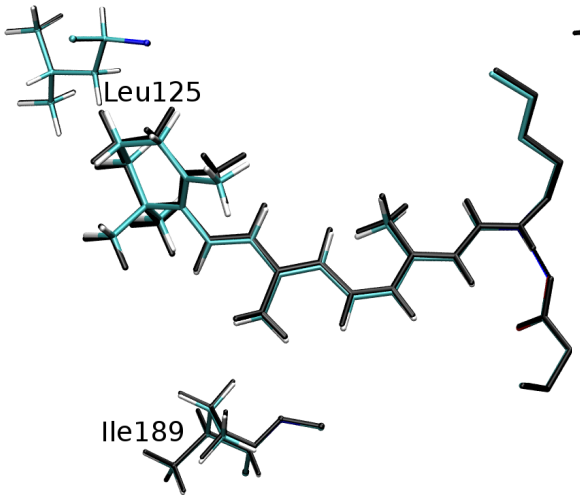
ment of more than 0.7 Å for C_5 (β-ionone ring) and C_{19} is a result of the last two mutations L125G and I189P (28fold → 30fold, Figure 45c).

Figure 46 and Table 33 show that both mutations have a significant influence on this displacement. The I189P mutation has a larger effect on the displacement of C_5 and C_{19} than L125G. This is hard to see in the figures but is quantified with vectorial projections of displacement vectors on the total displacement vector (28fold → 30fold). For the displacement of N_{16}, both mutations are equally responsible. Table 33 shows that the effect of a mutation X added to the 28fold mutand (28fold → 28fold+X) is larger than added to the 28fold+Y mutant (28fold+Y → 30fold).

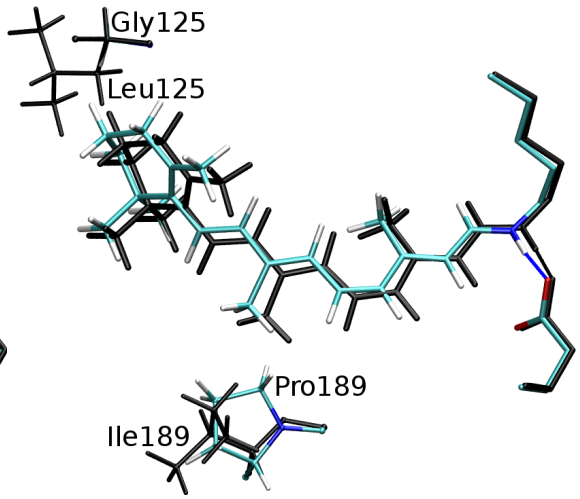
The effect of these two mutations on the excitation energy and the origin of this



(a) 24fold → 30fold mutant (24fold in black). Large displacement of retinal's π-system and the β-ionone ring.



(b) 24fold → 28fold mutant (24fold in black). Very limited difference in the β-ionone ring due to the L266V mutation. Leu125 and Ile189 are not mutated.



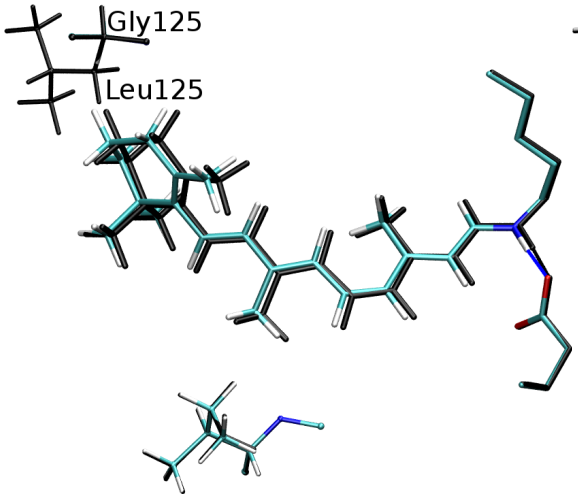
(c) 28fold → 30fold mutant (28fold in black). Same big difference in retinal's π-system and the β-ionone ring as between 24fold and 30fold mutants (Figure 45a).

Figure 45: Retinal displacement between the 24fold and 30fold mutant as a result of only two mutations (L125G and I189P).

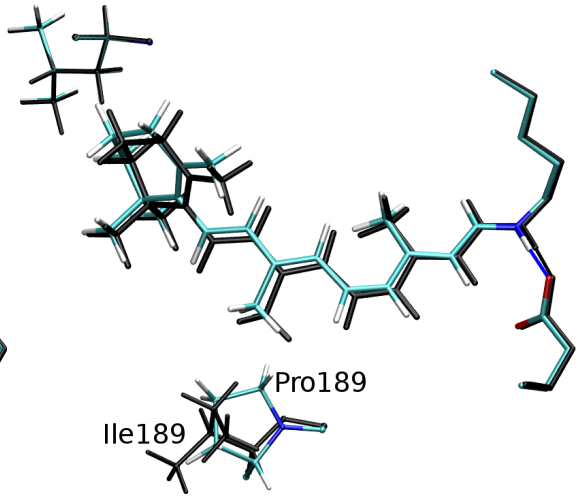
effect is described in detail in Section 5.4.2.4.

Table 33: Displacement of C_5 (β-ionone ring), C_{19} (methyl group at C_9), and N_{16} (PSB) between mutant structures (Å). The effects of the L125G and I189P mutations are given in % of the total displacement from 28fold to 30fold mutant.

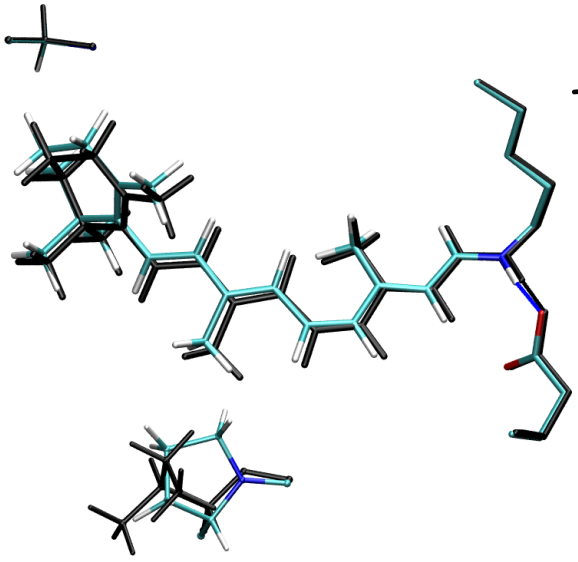
		C_5	C_{19}	N_{16}
24fold → 28fold		0.28	0.16	0.11
28fold → 30fold		0.70	0.74	0.37
Effect of L125G	28fold → 28+125	48%	41%	52%
	28+189 → 30fold	31%	26%	47%
Effect of I189P	28fold → 28+189	69%	74%	53%
	28+125 → 30fold	52%	59%	48%



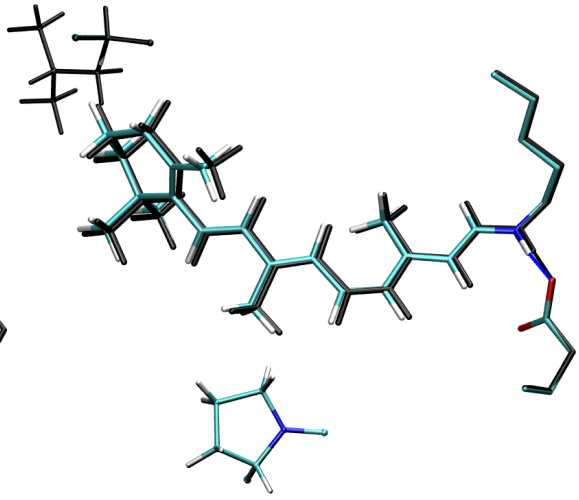
(a) 28fold → 28fold+125. Effect of the L125G mutation on the 28fold mutant.



(b) 28fold → 28fold+189. Effect of the I189P mutation on the 28fold mutant.



(c) 28fold+125 → 30fold. Effect of the I189P mutation on the 28fold+125 mutant.



(d) 28fold+189 → 30fold. Effect of the L125G mutation on the 28fold+189 mutant.

Figure 46: Detailed investigation on the origin of the retinal displacement between the 28fold and 30fold mutant. See Table 33 for a quantification of the displacements.

5.3.2.5 Summary of Retinal and Counterion Geometry

The distance of the PSB nitrogen to the hydrogen bonded $O_{\epsilon 2}$ (Table 34) is nearly unchanged in all multiple mutants. But the distance to the unbonded $O_{\epsilon 1}$ atom is largely shortened in the multiple mutants. This decreased distance is expected to have a hypsochromic effect on the excitation energy of the 10fold, 24fold, and 30fold mutants. The other distances in the HBN of the complex counterion are also typical bond lengths of hydrogen bonds (≈ 2.7 Å).

Table 34: Geometrical parameters (Å) of the PSB–counterion interaction (QM4 results).

	WT-Rh	10fold	24fold	30fold
$d(N_{PSB}–O_{\epsilon}1)$	3.66	3.54	3.32	3.20
$d(N_{PSB}–O_{\epsilon}2)$	2.70	2.75	2.73	2.71
$d(N_{PSB}–C_{\delta})$	3.59	3.56	3.40	3.34
$\varnothing(N_{PSB}–COO^{-})$	3.32	3.28	3.15	3.08
$d(O_{\epsilon}2–O_{Wat2b})$	2.78	2.71	2.70	2.69
$d(O_{\epsilon}2–O_{\gamma,Thr94})$	2.69	2.71	–	–

Figure 47 shows the bond lengths and dihedral angles of the six double and five single bonds in the retinal π-system in WT-Rh and in the 30fold mutant as an example for the three multiple mutants.

As Table 35 shows, the average of double bond lengths is decreased and the average of single bond length is increased in all multiple mutants (10fold, 24fold, 30fold). The bond-length alternation is increased from 6.69 10^{-2} Å to 7.91 10^{-2} Å. The average dihedral angle around single bonds is increased and the average dihedral angle around double bonds is decreased. This is a strong indication for a smaller conjugation of the π-system and a hypsochromic shift in excitation energy.

The bond-length alternation in retinal is a very good measure for the excitation energy. The larger the difference between the average single and double bond lengths, the smaller the conjugation of the π-system and the larger the excitation energy $\Delta E_{S_1 \leftarrow S_0}$. The linear correlation of the BLA and the excitation energy is shown in Section 5.4.2.5.

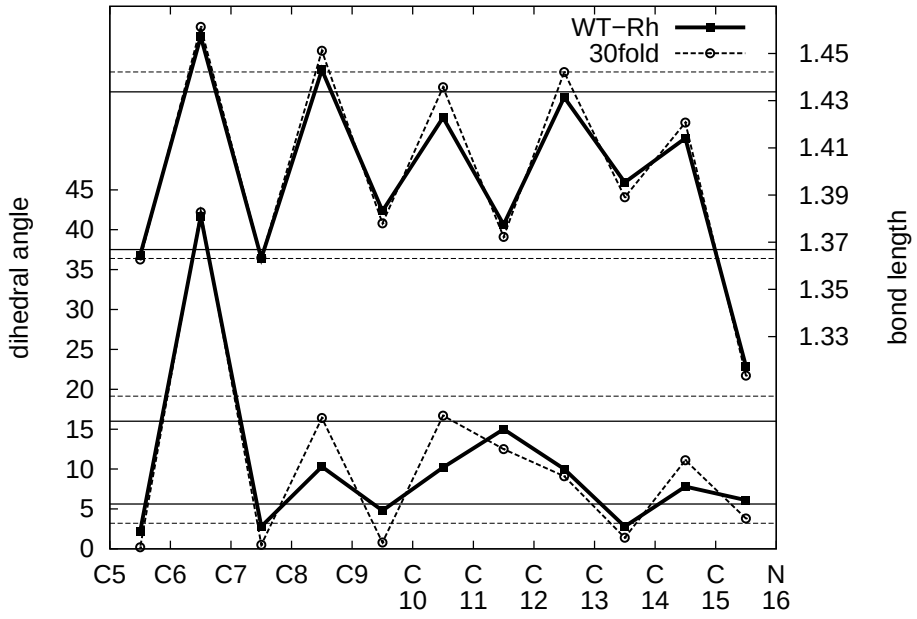


Figure 47: Bond lengths (top, Å) and dihedral angles (bottom, °) of the retinal π-system of WT-Rh and the 30fold mutant. Average values for single and double bonds are displayed as horizontal lines, separately.

Table 35: Average bond lengths (Å) and average dihedral angles (°) of single and double bonds in WT-Rh and the 30fold mutant.

		WT-Rh	10fold	24fold	30fold
bond	single	1.434	1.438	1.440	1.442
lengths	double	1.367	1.366	1.363	1.363
	BLA	$6.69\,10^{-2}$	$7.26\,10^{-2}$	$7.69\,10^{-2}$	$7.91\,10^{-2}$
dihedral	single	16.0	15.6	16.9	19.1
angles	double	5.6	4.11	3.96	3.2

5.4 Color Tuning Results

5.4.1 Single Mutants

For all structures of mutations described in Section 5.3.1 excitation energies were calculated with the SORCI method and CHARMM charges. The mutations are devided into three groups: Those investigated experimentally by Lin et al[117] (10, Section 5.4.1.1), those found by perturbation analysis (14, Section 5.4.1.2), and those found in an area of 5 Å around retinal and Lys296 side chain (6, Section 5.4.1.3).

5.4.1.1 Ten Mutations investigated by Lin et al.

The ten mutations investigated by Lin et al.[117] have very small electrostatic effects in WT-Rh except Trp265 and Glu122 because all other residues are apolar in WT-Rh (Met86, Val87, Gly90, Ala117, Ala124, Ala292, Ala295, Ala299). The residues being far away (>3 Å, M86L, V87C, A124T, A299C) experimentally have very small shifts in excitation energy and also the calculated values are small.

Table 36: Experimental and calculated shifts in excitation energy $\Delta\Delta E_{S_1 \leftarrow S_0}$ (SORCI(man), eV), electrostatic effect $ES^{\text{WT-Rh}}(I)$ (OM2/MRCI, eV), and minimal distance to retinal (Å).

	distance	$-ES^{\text{WT-Rh}}(I)$	exp.[a]	calc.
M86L	6.9	−0.001	–	0.00
V87C	10.2	0.000	0.01	0.02
G90S	6.3 (**2.7**[b])	—[c]	0.06	0.05 (0.05[d])
A117G	**2.8** (**2.3**[b])	0.003	0.02	−0.01
E122L[e]	**2.3**	−0.007	0.03	−0.02
A124T	5.6	0.000	0.02	0.03
W265Y	**2.1**	0.024	0.08	0.06 (0.09[d])
A292S	**3.0**	0.000	0.05	0.04
A295S_a	**2.8**	−0.002	0.03	−0.01
A295S_b[f]	dto.	dto.	dto.	0.05
A295S_c[f]	dto.	dto.	dto.	-0.06
A299C	9.1 (7.8[b])	0.000	0.01	0.00
sum		0.020	0.28	0.16

[a] From Lin et al.[117]
[b] Minimal distance to Wat2b.
[c] For Gly90 no electrostatic effect $ES^{\text{WT-Rh}}(I)$ can be calculated.
[d] Shift calculated with increased QM region (QM5_X) including the mutated residue X.
[e] Glu122 is protonated in WT-Rh. Otherwise E122L mutation would have a shift of ≈ 0.6 eV.[58]
[f] A295S_b is 0.9 kcal/mol higher in energy than A295S_a. A295S_c is 0.3 kcal/mol lower.

The deviation of the calculated shifts to the experimental ones are small (± 0.01 eV) for most of the mutations. The deviation of -0.02 eV for the W265Y mutation is acceptable and the inclusion of residue 265 in QM5_265 calculations of WT-Rh and the W265Y mutant gives a shift of 0.09 eV being closer to the experimental value.

A117G, E122L, and A295S mutations have larger errors (-0.03, -0.05, and -0.04 eV, respectively), all being calculated to be smaller. These mutations are very close to retinal and thus can have a large impact if differences in the surrounding are induced which were not observed in MD simulations. The large experimental shift of the E122L mutation is strange because the deletion of the Glu122 sidechain charges in WT-Rh gives a small bathochromic shift and thus a mutation into an apolar residue is expected to also have a bathochromic shift. All possible orientations of the Leu122 sidechain were tested but either they are instable in MD simulations, are higher in energy or have shifts in excitation energy being more negative. In 24fold and 30fold multiple mutants the orientation of Leu122 is changed. The A117G mutation is expected to be very simple because the glycine has no side chain atoms that can have different orientations but there seems to be a larger change in the surrounding protein that can not be modeled with MD simulations. The orientation of the Ser295 sidechain atoms were investigated in detail (see page 104).

The sum of shifts is 0.12 eV smaller than the experimental sum. Thus for multiple mutants the shifts in excitation energy are also expected to be too small.

The sum of electrostatic shifts is very far away from the sum of shifts in excitation energy. This is not surprising because the mutation types are mainly from apolar to polar or polar to polar (W265Y). The only polar to apolar mutation E122L has a small effect due to the orientation of Glu122 (the OH group being perpendicular to the last retinal double bond).

5.4.1.2 14 Mutations Found by Perturbation Analysis

For the 14 mutations found by perturbation analysis, no experimental shifts are known. Thus, the quality of the mutants can not be evaluated. But it can be seen in Table 37 that the shifts in excitation energies are large (sum of 0.25 eV) and thus these 14 mutations are expected to be very important for the overall shift between WT-Rh and HB.

Some of the shifts in excitation energy in the perturbation analysis and from the relaxed mutant structures are strongly different because the mutations either introduce new polar groups (Y191W, T297S) or have a structural effect on other residues. The sum of shifts of the mutants is the same if calculated with OM2/MRCI. Thus, the discrepancy is not because of the difference in the calculation method (OM2/MRCI, SORCI).

Table 37: Electrostatic effects of the 14 new residues selected in the perturbation analysis (OM2/MRCI, eV) and shifts in excitation energy $\Delta\Delta E_{S_1 \leftarrow S_0}$ of the 14 single mutants (SORCI(man),eV).

	$-ES^{\text{WT-Rh}}(I)$	$\Delta\Delta E_{S_1 \leftarrow S_0}$
Y43F	0.018	0.02
T92F	−0.004	0.02
T93P	0.006	0.03
T97A	0.007	0.01
N111A	0.008	0.03
C167V	0.012	0.04
Y178F	0.008	0.01
Y191W	0.026	0.06
Y206F	−0.004	−0.01
M207L	0.005	0.01
H211C	−0.014	−0.04
T297S	0.011	−0.01
S298A	−0.009	−0.02
sum	0.070	0.17
T94V	–	0.08

5.4.1.3 Six Mutations Found in 5 Å Around Retinal

The shifts in excitation energy of the six mutations additionally found in the binding pocket are very small. This is not surprising because they are seen to be electrostatically unimportant in perturbation theory. All mutations are into apolar AAs that are expected to have a small electrostatic effect. Again, the quality of these mutants can not be evaluated because of missing experimental shifts.

Table 38: Shifts in excitation energy $\Delta\Delta E_{S_1 \leftarrow S_0}$of the six single mutants (SORCI(man), eV).

L47V	−0.002
F91V	0.005
L125G	0.004
I189P	−0.003
L266V	−0.003
V300I	0.000
sum	0.001

The small sum of shifts would suggest a limited effect on the shift in multiple mutants but, as mentioned in Section 5.3.2.4, the structural effect of these mutations is large and also the shift in excitation energy of the multiple mutants is strongly effected as described in Section 5.4.2.4.

5.4.2 Multiple Mutants

5.4.2.1 Decomposition of Shifts in Excitation Energy

The shifts in excitation energy of single and multiple mutants can have several origins. For some of the multiple mutants the total shift is investigated in detail. Therefore, the electrostatic effects $ES^X(I)$ defined in Equation 31 are calculated and are compared with the electrostatic effects in WT-Rh $ES^{\mathrm{WT-Rh}}(I)$. The difference is called electrostatic shift:

$$\Delta ES^{\text{X–WT-Rh}}(I) = ES^X(I) - ES^{\mathrm{WT-Rh}}(I) \tag{32}$$

With this quantity, also the effect of changed orientations of conserved residues can be evaluated.

Another decomposition method of the shift in excitation energy is the assignment of shifts to the three important parts: (1) geometry of retinal $\Delta(\mathrm{Ret})$, (2) geometry of the complex counterion $\Delta(\mathrm{CCI})$, (3) electrostatic shift of the binding pocket $\Delta(\mathrm{bp})$. Therefore, the excitation energies with and without external charges (protein, vacuo) have to be calculated for WT-Rh and the mutant with different QM regions. To decompose $\Delta(\mathrm{CCI})$ into contributions from Glu113 ($\Delta(\mathrm{Glu113})$), Thr/Val94 and Wat2b ($\Delta(94, \mathrm{Wat2b})$), and Ser90 ($\Delta(\mathrm{Ser90})$), also calculations with QM2 and QM4 region are necessary.

The following definitions apply to a mutant with the GG90S mutation. Thus, the largest QM region for WT-Rh is QM4 and QM5_90 for the mutant. These regions can be compared directly because residue 90 is a glycine in WT-Rh.

$$\begin{aligned}\Delta(\mathrm{Ret}) &= \Delta\Delta E^{\text{X–WT-Rh}}_{\text{vacuo}}(\mathrm{QM1}) \\ &= \Delta E^{X}_{\text{vacuo}}(\mathrm{QM1}) - \Delta E^{\text{WT-Rh}}_{\text{vacuo}}(\mathrm{QM1})\end{aligned} \tag{33}$$

$$\Delta(\mathrm{Glu113}) = \Delta\Delta E^{\text{X–WT-Rh}}_{\text{vacuo}}(\mathrm{QM2}) - \Delta(\mathrm{Ret}) \tag{34}$$

$$\begin{aligned}\Delta(94, \mathrm{Wat2b}) &= \Delta\Delta E^{\text{X–WT-Rh}}_{\text{vacuo}}(\mathrm{QM4}) - \Delta\Delta E^{\text{X–WT-Rh}}_{\text{vacuo}}(\mathrm{QM2}) \\ &= \Delta\Delta E^{\text{X–WT-Rh}}_{\text{vacuo}}(\mathrm{QM4}) - \Big(\Delta(\mathrm{Glu113}) + \Delta(\mathrm{Ret})\Big)\end{aligned} \tag{35}$$

$$\begin{aligned}\Delta(\mathrm{Ser90}) &= \Delta\Delta E^{\text{X–WT-Rh}}_{\text{vacuo}}(\mathrm{QM5\_90}) - \Delta\Delta E^{\text{X–WT-Rh}}_{\text{vacuo}}(\mathrm{QM4}) \\ &= \Delta\Delta E^{\text{X–WT-Rh}}_{\text{vacuo}}(\mathrm{QM5\_90}) - \Big(\Delta(94, \mathrm{Wat2b}) + \Delta(\mathrm{Glu113}) + \Delta(\mathrm{Ret})\Big)\end{aligned} \tag{36}$$

These last three shifts $\Delta(\mathrm{Glu113})$, $\Delta(94, \mathrm{Wat2b})$, and $\Delta(\mathrm{Ser90})$ can be combined to the shift of the complex counterion

$$\begin{aligned}\Delta(\mathrm{CCI}) &= \Delta(\mathrm{Glu113}) + \Delta(94, \mathrm{Wat2b}) + \Delta(\mathrm{Ser90}) \\ &= \Delta\Delta E^{\text{X–WT-Rh}}_{\text{vacuo}}(\mathrm{QM5}) - \Delta(\mathrm{Ret}).\end{aligned} \tag{37}$$

The shift of the binding pocket is just

$$\begin{aligned}\Delta(\mathrm{bp}) &= \Delta\Delta E^{\text{X–WT-Rh}}_{\text{protein}}(\mathrm{QM5\_90}) - \Delta\Delta E^{\text{X–WT-Rh}}_{\text{vacuo}}(\mathrm{QM5\_90}) \\ &= \Delta\Delta E^{\text{X–WT-Rh}}_{\text{protein}}(\mathrm{QM5\_90}) - \Big(\Delta(\mathrm{Ret}) + \Delta(\mathrm{CCI})\Big).\end{aligned} \tag{38}$$

5.4.2.2 Multiple Mutants by Lin et al.

The shifts in excitation energy of the multiple mutants investigated by Lin et al.[117] are given in Table 39.

Table 39: Experimental and calculated shifts in excitation energy $\Delta\Delta E_{S_1 \leftarrow S_0}$ (SORCI, eV) of multiple mutants and sum of shifts of the single mutants (differences in brackets).

	exp. (from Ref. 117)		calc.	
Mutant	$\Delta\Delta E_{S_1 \leftarrow S_0}$	sum[a]	$\Delta\Delta E_{S_1 \leftarrow S_0}$	sum[a]
h2	0.07 (−0.01)[b]	0.06	0.03 (−0.02)	0.05
h2′	0.06 (−0.01)	0.07	0.04 (−0.03)	0.07
h3	0.06 (±0.00)	0.06	0.03 (+0.03)	0.00
h7	0.08 (±0.00)	0.08	0.04 (+0.02)	0.02
h27	0.13 (−0.01)	0.14	0.05 (−0.02)	0.07
h36	0.15 (+0.01)	0.14	0.13 (+0.06)	0.07
h237	0.14 (−0.06)	0.20	0.09 (+0.01)	0.08
h267	0.27 (+0.06)	0.21	0.15 (+0.01)	0.14
h267+117	0.27 (+0.04)	0.23	0.17 (+0.04)	0.13
h267+122	0.33 (+0.07)	0.24	0.15 (+0.03)	0.12
h267+124	0.20 (−0.03)	0.23	0.17 (±0.00)	0.17
9fold	0.35 (+0.08)	0.27	0.21 (+0.07)	0.14
10fold	0.36 (+0.08)	0.28	0.22 (+0.06)	0.16

[a] Sum of the shifts of the single mutants included in the multiple mutant.

[b] Difference to the sum of shifts of the single mutants.

Most of the calculated shifts are smaller than the experimental ones. Trusting the calculation method, this might have two possible reasons: Either the single mutants are structurally wrong and these errors sum up to a larger error in the multiple mutants, or some cooperative effects have not been found that would produce the missing shift.

Cooperative effects play a key role at these multiple mutants. This can be seen by the large differences between the shift and the sum of shifts of the single mutants.

There is hardly a trend observable in these cooperative effects, neither in the experimental results nor in the calculated ones. This is because a cooperative effect, e.g. in the h267 multiple mutant, can have a hypsochromic shift. But if the A124T mutation is added, the structural change is reversed and thus no cooperative effect is observable in the h267+124 multiple mutant. Cooperative effects can either be movements of helices induced by some mutations or changes in the orientation of some residues that are mutated or even conserved. The first effects can not be modeled with the methodology in this work because of restrained positions of many atoms in the system. Only local changes can be observed in optimizations. Non-local changes can be modeled in MD simulations if the

barriers of the movements are small enough. For large barriers, these changes can only be built by chemical intuition or by luck. For some multiple mutants non-local changes in HBNs or in residue orientations were found. These were described in Section 5.3.2.

An example for a structural cooperative effect is the complex counterion that is changed due to the new water Wat46 in the h267 multiple mutant. Ser90 is differently oriented than in the h27 multiple mutant and thus the shift from h27 to h267 (0.14 and 0.10 eV experimentally calculated) is very large compared to the shift of the W265Y single mutant (0.08 and 0.06 eV).

Color Tuning in the 10fold Mutant

In the same manner as the electrostatic effects have been calculated for WT-Rh, this can be done for the 10fold mutant to get electrostatic shifts of all residues. Table 40 shows the electrostatic effects and shifts. Also some conserved residues slightly change their orientation or distance to retinal and thus have different electrostatic effects in WT-Rh and in the 10fold mutant. These are shown in Table 41. The sums of 0.18 and 0.04 eV give a total hypsochromic shift of 0.22 eV which is exactly the shift of the 10fold mutant. This concurrence does not mean that only the electrostatic effects are important in color tuning. The effects that play a role are complex and can only be approximately assigned with values.

The decomposition of the total shift into parts for retinal geometry, counterion

Table 40: Electrostatic effects of mutated residues in the 10fold mutant and electrostatic shifts relative to WT-Rh (OM2/MRCI, eV).

Residue	$ES^{\text{WT-Rh}}$	ES^{10fold}	$\Delta ES^{\text{10fold–WT-Rh}}$	Mutation
86	0	0	0	Met→Leu
87	0	0	0	Val→Cys
90	—[a]	−0.02	−0.02	Gly→Ser
117	0	—[a]	0	Ala→Gly
122	0.01	0	−0.01	Glu→Leu
124	0	0.01	0.01	Ala→Thr
265	−0.02	−0.01	0.01	Trp→Tyr
292	0	0.06	0.06	Ala→Ser
295	0	0.02	0.02	Ala→Ser
299	0	0	0	Ala→Cys
Wat45	—[b]	0.09	0.09	new water
Wat46	—[b]	0.02	0.02	new water
sum			0.18	

[a] For Gly90 in WT-Rh and Gly117 in the 10fold mutant, no electrostatic effect can be calculated.

[b] Wat45 and Wat46 are newly introduced in the 10fold mutant.

Table 41: Electrostatic effects of some conserved residues in the 10fold mutant and electrostatic shifts relative to WT-Rh (OM2/MRCI, eV).

Residue		$ES^{\text{WT-Rh}}$	ES^{10fold}	$\Delta ES^{\text{10fold–WT-Rh}}$
Tyr	43	−0.02	−0.01	0.01
Met	44	0.01	0	0
Glu	113[a]	0.46	0.51	0.05
Thr	118	−0.05	−0.03	0.01
Glu	181	0.06	0.05	−0.01
Ser	186	−0.03	−0.05	−0.02
Tyr	268	−0.02	−0.02	0
Phe	293	0.02	0.01	−0.01
water	6	−0.01	−0.01	0.01
water	Wat2a	0.05	0.06	0.01
water	14	−0.01	0.01	0.01
water	Wat2b[b]	−0.01	−0.02	−0.01
sum				0.04

[a] Glu113 is part of the QM4 and QM2 region. Thus, the electrostatic effect was calculated with the small QM1 region.
[b] Wat2b is part of the QM4 region. Thus, the electrostatic effect was calculated with the medium sized QM2 region.

Table 42: Shifts in excitation energy between WT-Rh and the 10fold mutant with and without external charges (SORCI(nat19), eV).

	$\Delta\Delta E^{\text{10fold–WT-Rh}}_{\text{vacuo}}$	$\Delta\Delta E^{\text{10fold–WT-Rh}}_{\text{protein}}$
QM5_90	−0.01	0.23
QM4	−0.02	0.22
QM2	0.04	0.19
QM1	−0.01	0.20

effects, and electrostatic effects of the binding pocket is done with Equations 33 to 38. The resulting shifts for all QM regions calculated with SORCI(NAT19) and the single shifts are given in Tables 42 and 43.

The large shift of the counterion Glu113 is because of the closer distance to the PSB in the 10fold mutant (see Section 5.3.2.5). This effect is counterbalanced by the changed geometry of the complex counterion. Thus, the shift from the binding pocket (0.24 eV) is the most important part.

Table 43: Decomposition of the shift between WT-Rh and the 10fold mutant (eV).

Δ(Ret)	-0.01	-4%
Δ(CCI)	0.00	0%
Δ(Glu113)	0.05	
Δ(94, Wat2a)	-0.06	
Δ(Ser90)	0.01	
Δ(bp)	0.24	104%
$\Delta\Delta E_{\text{protein}}^{\text{10fold–WT-Rh}}$(QM5_90)	0.23	100%

5.4.2.3 24fold Mutants

Table 44 shows the results for the multiple mutants with 24 mutations. The instable structure 24fold_b with a third newly added water (described in Section 5.3.2.2) gives a very small shift in excitation energy, only 0.03 eV higher than the 10fold mutant. This is because of the changed HBN, especially the changed orientation of Wat46.

Table 44: Shifts in excitation energy $\Delta\Delta E_{S_1 \leftarrow S_0}$ of multiple mutants (SORCI(man), eV).

10fold	0.22
sum of 14	0.25
24fold	0.36
24fold_b[a]	*0.25*

[a] 24fold mutant with a third newly added water and a strongly changed HBN of Tyr265, waters 45–47, Wat2b and Ser90.

On the structural level, the 24fold mutant shows cooperative effects, i.e. the structure is not simply the overlay of the single mutants. In the excitation energy these cooperative effects can be observed, too. The 24fold mutant has a shift in excitation energy of 0.36 eV, being much smaller than the sum of the 10fold mutant and the 14 new mutations ($0.22 + 0.25 = 0.47$ eV).

The color tuning in the 24fold mutant can be described as the shift from WT-Rh or from the 10fold mutant. The shift from 10fold is investigated in Section 5.4.2.5. The mutated residues and the two newly induced water molecules give a sum of electrostatic shifts of 0.26 eV. A large part results from the 14 new mutations. Most of them have negative $ES^{\text{WT}-\text{Rh}}$ and are mutated in apolar residues with no ES^{24fold}, resulting in a positive $\Delta ES^{\text{24fold–WT-Rh}}$. The sum of $\Delta ES^{\text{24fold–WT-Rh}}$ of the conserved residues (0.06 eV) is similar to the 10fold mutant (0.04 eV)

Table 45: Electrostatic effects of mutated residues in the 24fold mutant and electrostatic shifts relative to WT-Rh (OM2/MRCI, eV).

Residue	$ES^{\text{WT-Rh}}$	ES^{24fold}	$\Delta ES^{\text{24fold–WT-Rh}}$	Mutation
86	0	0	0	Met→Leu
87	0	0	0	Val→Cys
90	—[a]	−0.02	−0.02	Gly→Ser
117	0	—[a]	0	Ala→Gly
122	0.01	0	−0.01	Glu→Leu
124	0	0.01	0.01	Ala→Thr
265	−0.02	0.01	0.02	Trp→Tyr
292	0	0.05	0.05	Ala→Ser
295	0	0.01	0.01	Ala→Ser
299	0	0	0	Ala→Cys
Wat45	—[b]	0.06	0.06	new water
Wat46	—[b]	0.03	0.03	new water
43	−0.02	0	0.02	Tyr→Phe
92	0	0	0	Thr→Phe
93	−0.01	0	0.01	Thr→Pro
94[c]	−0.02	0	0.02	Thr→Val
97	−0.01	0	0.01	Thr→Ala
111	−0.01	0	0.01	Asn→Ala
167	−0.01	0	0.01	Cys→Val
178	−0.01	0	0.01	Tyr→Phe
191	−0.03	0	0.02	Tyr→Trp
206	0	0	0	Tyr→Phe
207	0	0	0	Met→Leu
211	0.01	0	−0.01	His→Cys
297	−0.01	0.01	0.02	Thr→Ser
298	0.01	−0.01	−0.02	Ser→Ala
sum			0.26	

[a] For Gly90 in WT-Rh and Gly117 in the 24fold mutant, no electrostatic effect can be calculated.
[b] Wat45 and Wat46 are newly introduced.
[c] Thr/Val94 is part of the QM4 region. Thus, the electrostatic effect was calculated with the medium sized QM2 region.

Table 46: Electrostatic effects of some conserved residues in the 24fold mutant and electrostatic shifts relative to WT-Rh (OM2/MRCI, eV).

Residue		$ES^{\text{WT-Rh}}$	ES^{24fold}	$\Delta ES^{\text{24fold–WT-Rh}}$
Met	44	0.01	0	0
Glu	113[a]	0.46	0.54	0.08
Thr	118	−0.05	−0.03	0.02
Glu	181	0.06	0.03	−0.03
Ser	186	−0.03	−0.05	−0.02
Leu	266	0	−0.01	−0.01
Tyr	268	−0.02	−0.01	0.01
Phe	293	0.02	0.01	−0.01
water	6	−0.01	−0.01	0.01
water	Wat2a	0.05	0.06	0.01
water	14	−0.01	0	0.01
water	Wat2b[b]	−0.01	−0.02	0
sum				0.06

[a] Glu113 is part of the QM4 and QM2 region. Thus, the electrostatic effect was calculated with the small QM1 region.
[b] Wat2b is part of the QM4 region. Thus, the electrostatic effect was calculated with the medium sized QM2 region.

Tables 47 and 48 show the calculated shifts between WT-Rh and the 24fold mutant with different QM regions, with and without external charges, and the decomposition of the shift into the three parts. The hypsochromic shift of the counterion ($\Delta(\text{Glu113}) = 0.06\,\text{eV}$) is not counterbalanced by the other residues in the complex counterion because of the T94V mutation and thus $\Delta(\text{CCI})$ is hypsochromic, too. Again the shift of the binding pocket $\Delta(\text{bp})$ is the largest part of the total 0.37 eV shift.

Table 47: Shifts in excitation energy between WT-Rh and the 24fold mutant with and without external charges (SORCI(nat19), eV).

	$\Delta\Delta E^{\text{24fold–WT-Rh}}_{\text{vacuo}}$	$\Delta\Delta E^{\text{24fold–WT-Rh}}_{\text{protein}}$
QM5_90	0.03	0.37
QM4	0.06	0.35
QM2	0.04	0.32
QM1	−0.02	0.32

Table 48: Decomposition of the shift between WT-Rh and the 24fold mutant (eV).

Δ(Ret)	-0.02	-5%
Δ(CCI)	0.05	14%
Δ(Glu113)	0.06	
Δ(94, Wat2a)	0.02	
Δ(Ser90)	-0.03	
Δ(bp)	0.34	92%
$\Delta\Delta E_{\text{protein}}^{\text{24fold–WT-Rh}}$(QM5_90)	0.37	100%

5.4.2.4 30fold Mutant

For the 30fold mutant cooperative effects are very important, too. The shifts of the six new mutations cancel each other but the 30fold mutant has a shift being 0.06 eV higher than the 24fold mutant. The origin of this large shift was mentioned to be the L125G and I189P mutations that are responsible for a large translocation of retinal towards the β-ionone ring (described in Section 5.3.2.4).

The 28fold mutant (without L125G and I189P) has nearly the same excitation energy as the 24fold mutant, but the two different 28fold+X mutants are $\approx$ 0.03 eV higher and together they give the full 0.06 eV shift to the 30fold mutant. This surprisingly high effect is not because of electrostatics. All electrostatic effects $ES^{\text{WT-Rh}}(I)$ and $ES^{\text{30fold}}(I)$ and thus also both electrostatic shifts $\Delta ES^{\text{30fold-WT-Rh}}(I)$ are smaller than 0.01 eV.

The negatively charged Glu181 is located close to the middle of retinal and does not change its position. In total, Glu181 is closer to the PSB in the 30fold mutant and further away from the β-ionone ring, resulting in a higher shift in excitation energy. Table 49 shows the movements and the shifts in excitation energy.

The electrostatic shifts of the mutated residues are very similar to the shifts in the 24fold mutant (see Section 5.4.2.5). The six new mutations all have a

Table 49: Shifts in excitation energy of multiple mutants (SORCI(man), eV) and distances from Glu181$O_{\epsilon 1}$ to N_{16} (PSB) and C_6 (β-ionone ring) (Å).

	$\Delta\Delta E_{S_1 \leftarrow S_0}$	Glu181 $O_{\epsilon 1}$ to N_{16}	C_6
24fold	0.36	7.3	9.0
28fold	0.35	7.3	9.0
29fold_125	0.39	7.2	9.1
29fold_189	0.40	7.2	9.1
30fold	0.42	7.1	9.2

$\Delta ES^{\text{30fold–WT-Rh}} < 0.005\,\text{eV}$. Thus, the sum of electrostatic shifts (0.24 eV) is also similar to 24fold (0.26 eV). The electrostatic shifts of Glu113 and Wat2b are larger making the sum of $\Delta ES^{\text{30fold–WT-Rh}}$ of the conserved residues larger (0.12 eV).

Table 50: Electrostatic effects of mutated residues in the 30fold mutant and electrostatic shifts relative to WT-Rh (OM2/MRCI, eV)[a].

Residue	$ES^{\text{WT-Rh}}$	ES^{30fold}	$\Delta ES^{\text{30fold–WT-Rh}}$	Mutation
86	0	0	0	Met→Leu
87	0	0	0	Val→Cys
90	—[b]	−0.02	−0.02	Gly→Ser
117	0	—[b]	0	Ala→Gly
122	0.01	0	−0.01	Glu→Leu
124	0	0.01	0.01	Ala→Thr
265	−0.02	−0.01	0.02	Trp→Tyr
292	0	0.05	0.05	Ala→Ser
295	0	0.01	0.01	Ala→Ser
299	0	0	0	Ala→Cys
Wat45	—[c]	0.06	0.06	new water
Wat46	—[c]	0.03	0.03	new water
43	−0.02	0	0.02	Tyr→Phe
92	0	0	0	Thr→Phe
93	−0.01	0	0.01	Thr→Pro
94[d]	−0.02	0	0.02	Thr→Val
97	−0.01	0	0.01	Thr→Ala
111	−0.01	0	0.01	Asn→Ala
167	−0.01	0	0.01	Cys→Val
178	−0.01	0	0.01	Tyr→Phe
191	−0.03	0	0.03	Tyr→Trp
206	0	0	0	Tyr→Phe
207	0	0	0.01	Met→Leu
211	0.01	0	−0.01	His→Cys
297	−0.01	−0.01	0	Thr→Ser
298	0.01	0	−0.01	Ser→Ala
sum			0.24	

[a] Electrostatic effects and electrostatic shifts of the six new mutations are all $< 0.005\,\text{eV}$.
[b] For Gly90 in WT-Rh and Gly117, Gly125 in the 30fold mutant, no electrostatic effect can be calculated.
[c] Wat45 and Wat46 are newly introduced.
[d] Thr/Val94 is part of the QM4 region. Thus, the electrostatic effect was calculated with the medium sized QM2 region.

Table 51: Electrostatic effects of some conserved residues in the 30fold mutant and electrostatic shifts relative to WT-Rh (OM2/MRCI, eV).

Residue		$ES^{\text{WT-Rh}}$	ES^{30fold}	$\Delta ES^{\text{30fold–WT-Rh}}$
Met	44	0.01	0	−0.01
Glu	113[a]	0.46	0.56	0.10
Thr	118	−0.05	−0.02	0.02
Glu	181	0.06	0.05	−0.02
Ser	186	−0.03	−0.05	−0.02
Tyr	268	−0.02	−0.01	0.02
Phe	293	0.02	0.01	−0.01
water	6	−0.01	−0.01	0.01
water	Wat2a	0.05	0.06	0.01
water	14	−0.01	0.01	0.01
water	Wat2b[b]	−0.01	−0.02	−0.01
sum				0.12

[a] Glu113 is part of the QM4 and QM2 region. Thus, the electrostatic effect was calculated with the small QM1 region.
[b] Wat2b is part of the QM4 region. Thus, the electrostatic effect was calculated with the medium sized QM2 region.

Tables 52 and 53 show the calculated shifts between WT-Rh and the 30fold mutand with different QM regions, with and without external charges, and the decomposition of the shift into the three parts. The large hypsochromic shift of the counterion ($\Delta(\text{Glu113}) = 0.11\,\text{eV}$) is not counterbalanced by the other residues in the complex counterion giving a large $\Delta(\text{CCI})$ of 0.06 eV. Again, the shift of the binding pocket $\Delta(\text{bp})$ is the largest part of the total 0.42 eV shift.

Table 52: Shifts in excitation energy between WT-Rh and the 30fold mutant with and without external charges (SORCI(nat19), eV).

	$\Delta\Delta E^{\text{30fold–WT-Rh}}_{\text{vacuo}}$	$\Delta\Delta E^{\text{30fold–WT-Rh}}_{\text{protein}}$
QM5_90	0.08	0.42
QM4	0.11	0.38
QM2	0.13	0.39
QM1	0.02	0.41

Table 53: Decomposition of the shift between WT-Rh and the 30fold mutant (eV).

$\Delta(\text{Ret})$	0.02	5%
$\Delta(\text{CCI})$	0.06	14%
$\Delta(\text{Glu113})$	0.11	
$\Delta(94, \text{Wat2a})$	0.02	
$\Delta(\text{Ser90})$	-0.03	
$\Delta(\text{bp})$	0.34	81%
$\Delta\Delta E^{\text{30fold–WT-Rh}}_{\text{protein}}(\text{QM5\_90})$	0.42	100%

5.4.2.5 Summary of Multiple Mutant Results

The effect of the changed retinal geometry is very limited, only 0.02 eV between WT-Rh and 30fold mutant in the QM1 calculations without external charges. This is also a hint for the statement that the increased BLA in the 30fold mutant is not the reason for the large hypsochromic shift but a result of the large shift. Geometry optimizations were always done with external charges and thus the geometry without the changed electrostatics can not be calculated.

Table 54: Decomposition of the shift between WT-Rh and the multiple mutants (eV).

	10fold	24fold	30fold
$\Delta(\text{Ret})$	-0.01 (-4%)	-0.02 (-5%)	0.02 (5%)
$\Delta(\text{CCI})$	0.00 (0%)	0.05 (14%)	0.06 (14%)
$\Delta(\text{Glu113})$	0.05	0.06	0.11
$\Delta(94, \text{Wat2a})$	-0.06	0.02	0.02
$\Delta(\text{Ser90})$	0.01	-0.03	-0.03
$\Delta(\text{bp})$	0.24 (104%)	0.34 (92%)	0.34 (81%)
$\Delta\Delta E^{\text{X–WT-Rh}}_{\text{protein}}(\text{QM5\_90})$	0.23 (100%)	0.37 (100%)	0.42 (100%)

$\Delta(\text{Glu113})$ is increasing consistent with the decrease in the distance of PSB and Glu113 $\text{COO}^{\ominus}$-group (see Table 34) and with the increasing $ES^{\text{X}}(\text{Glu113})$. $\Delta(94, \text{Wat2a})$ is much smaller in the 24fold and 30fold mutants because of the T94V mutation and the negative $ES^{\text{WT}-\text{Rh}}(\text{Thr94}) = -0.02\,\text{eV}$. The by far largest shift comes from the changed electrostatics in the binding pocket ($\Delta(\text{bp})$), being 81% of the total shift in the 30fold mutant.

In Tables 55 and 56 the electrostatic shifts between WT-Rh, 10fold, 24fold, and 30fold mutant are given. The electrostatic effects of the first ten mutations do not change in the 24fold and 30fold mutant notably. Only the water molecules Wat45 and Wat46 give a shift due to the change in the complex counterion that is observable in the large numbers for residue 94, Glu113 and Wat2b. The electro-

static effects of the next 14 mutations do not change largely between the 24fold and 30fold mutant because most of them are mutated into apolar residues and do not have a different effect, even if the orientation is changed due to some cooperative effects. The last six mutations have electrostatic effects being close to zero in all models and thus have no electrostatic shifts. The largest changes in Table 56 are in the first column. The later mutations do not change these residues notably except the complex counterion residues (Glu113, Wa2b) and Glu181 that is responsible for a part of the hypsochromic shift from 24fold to 30fold mutant due to the large movement of retinal described in Section 5.3.2.4.

With different QM regions the BLA is slightly different but especially for QM2, QM4 and QM5_90 it is very similar. BLA with QM1 is smaller by 0.1 to 0.2 10^{-2} Å. Table 57 shows the shifts in excitation energy for the multiple mutants with different QM regions.

Plotting the excitation energies against the BLA gives a strong linear correlation of the excitation energy with the BLA (Figure 48) which is not the case for the excitation energies in vacuo (Figure 49). Especially for QM1 without charges, the excitation energy differs by ± 0.02 eV, although here the largest effect of the pure BLA is expected. This shows that the change in BLA is only a product of the bathochromic shift and not vice versa.

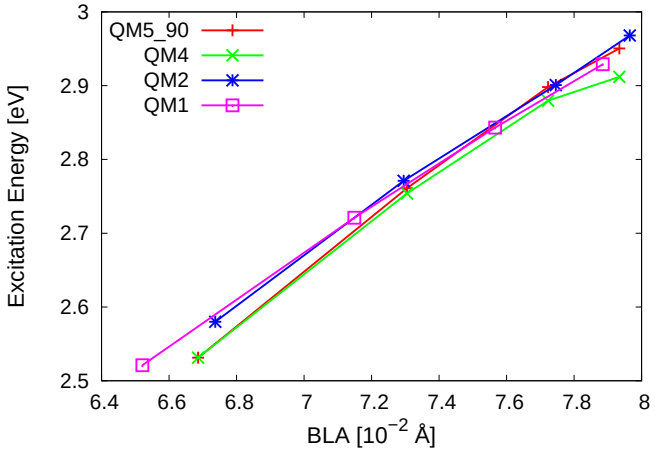


Figure 48: Linear correlation of BLA and excitation energy for all QM regions with external charges (protein). Regression equation is $\Delta E \approx 0.3 \cdot \mathrm{BLA} + 0.5$

Plotting the excitation energies of all single and multiple mutants against the BLA gives Figure 50. The linear dependence is the same as in the work of Hoffmann et al.[43] for archael rhodopsins (0.34 eV/10^{-2} Å). The vertical offset is because of the different chromophore (all-*trans* retinal in bR).

For WT-Rh and the 30fold mutants the MD simulations were performed with the small QM1 region. Snapshots from the trajectory of the first 520 ps are saved every ps and the excitation energy is calculated with the fast OM2/MRCI without structural relaxation.

Plotting the excitation energies of all these snapshots against their BLA also gives a good correlation. The linear slope of the correlation (0.06 and 0.08 eV/10^{-2} Å) is much smaller than in Figure 50 (see Table 57). Hence, this is another argument

Table 55: Electrostatic shifts of mutated residues between WT-Rh and the multiple mutants (OM2/MRCI, eV).

Residue	$\Delta ES^{10\text{fold–WT-Rh}}$	$\Delta ES^{24\text{fold–10fold}}$	$\Delta ES^{30\text{fold–24fold}}$	Mutation
86	0	0	0	Met→Leu
87	0	0	0	Val→Cys
90	−0.02	0	0	Gly→Ser
117	0	0	0	Ala→Gly
122	−0.01	0	0	Glu→Leu
124	0.01	0	0	Ala→Thr
265	0.01	0	0	Trp→Tyr
292	0.06	0	−0.01	Ala→Ser
295	0.02	−0.01	0	Ala→Ser
299	0	0	0	Ala→Cys
wat45	0.09	−0.03	0	new water
wat46	0.02	0.01	−0.01	new water
43	0.01	0.01	0	Tyr→Phe
92	0	0	0	Thr→Phe
93	0	0.01	0	Thr→Pro
94[a]	0	0.02	0	Thr→Val
97	0	0	0	Thr→Ala
111	0	0.01	0	Asn→Ala
167	0	0.01	0	Cys→Val
178	0	0.01	0	Tyr→Phe
191	0	0.02	0	Tyr→Trp
206	0	0	0	Tyr→Phe
207	0	0	0	Met→Leu
211	0	−0.01	0	His→Cys
297	0	0.02	−0.02	Thr→Ser
298	0	−0.01	0.01	Ser→Ala
266[b]	0	−0.01	0.01	Leu→Val
sum	0.19	0.06	0.00	

[a] Thr/Val94 is part of the QM4 region. Thus, the electrostatic effect was calculated with the medium sized QM2 region.

[b] The other five mutations all have electrostatic effects and electrostatic shifts <0.005 eV.

Table 56: Electrostatic shifts of some conserved residues between WT-Rh and the multiple mutants (OM2/MRCI, eV).

Residue		$\Delta ES^{10fold–WT-Rh}$	$\Delta ES^{24fold–10fold}$	$\Delta ES^{30fold–24fold}$
Met	44	0	0	0
Glu	113[a]	0.05	0.04	0.02
Thr	118	0.01	0	0
Glu	181	−0.01	−0.02	0.02
Ser	186	−0.02	0	0
Tyr	268	0	0.01	0.01
Phe	293	−0.01	0	0
water	6	0.01	0	0
water	Wat2a	0.01	0	0
water	14	0.01	0	0
water	Wat2b[b]	−0.01	0.01	0
sum		0.03	0.04	0.04

[a] Glu113 is part of the QM4 and QM2 region. Thus, the electrostatic effect was calculated with the small QM1 region.
[b] Wat2b is part of the QM4 region. Thus, the electrostatic effect was calculated with the medium sized QM2 region.

Table 57: BLA (10^{-2} Å), $\Delta E_{S_1 \leftarrow S_0}$ of WT-Rh and $\Delta\Delta E_{S_1 \leftarrow S_0}$ of multiple mutants with (protein) and without all external charges (vacuo) (SORCI(nat19), eV). Slope of excitation energy against BLA (eV/10^{-2} Å)

		WT-Rh	10fold	24fold	30fold	slope
QM5_90	BLA	6.69 [a]	7.30	7.72	7.93	
	vacuo	2.76 [a]	-0.01	0.03	0.08	0.06
	protein	2.53 [a]	0.23	0.37	0.42	0.34
QM4	BLA	6.69	7.26	7.69	7.91	
	vacuo	2.76	-0.02	0.06	0.11	0.08
	protein	2.53	0.22	0.35	0.38	0.31
QM2	BLA	6.74	7.30	7.75	7.97	
	vacuo	2.79	0.04	0.04	0.13	0.09
	protein	2.58	0.19	0.32	0.39	0.31
QM1	BLA	6.52	7.15	7.57	7.89	
	vacuo	1.98	-0.01	-0.02	0.02	0.01
	protein	2.52	0.20	0.32	0.41	0.30

[a] For WT-Rh no QM5_90 calculations were performed because residue 90 is a glycine and the QM4 results are taken instead.

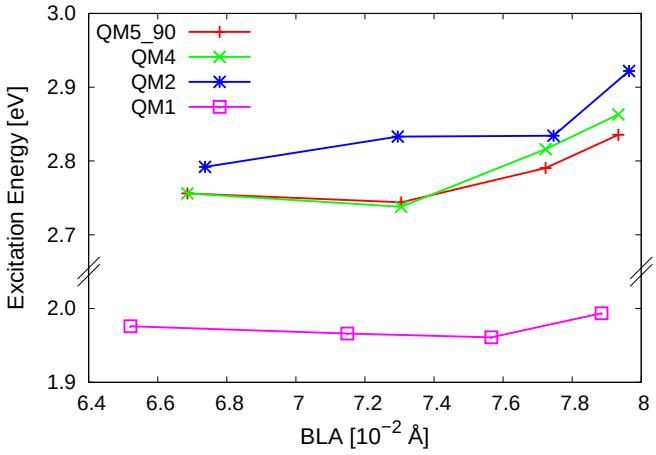


Figure 49: Excitation energy without external charges (vacuo) for all QM regions against BLA.

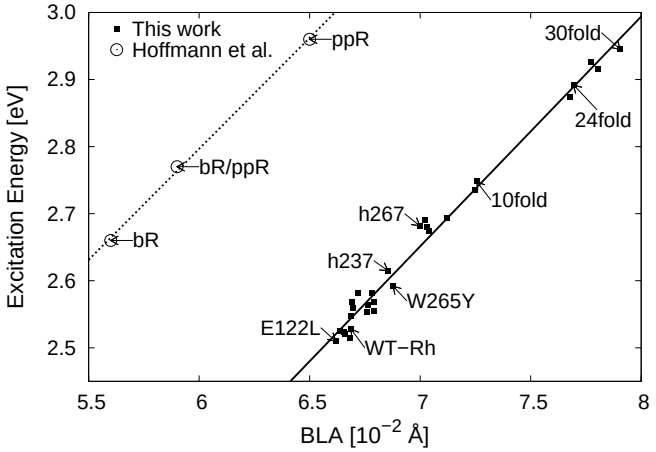


Figure 50: Correlation of BLA and excitation energy for all single and multiple mutants (SORCI(man), QM4, eV). Dotted: Results by Hoffmann et al.[43]

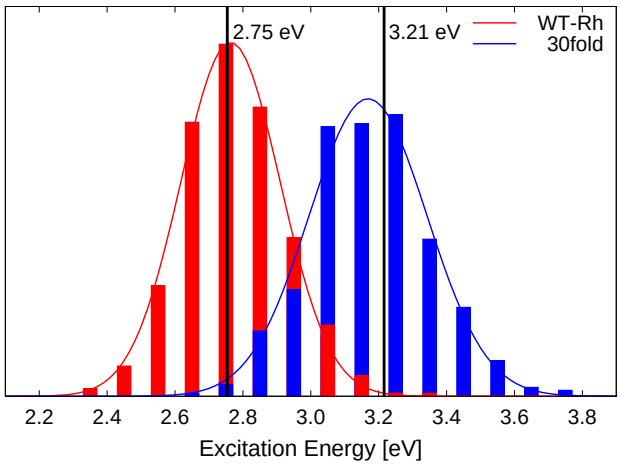


Figure 51: Histogram of $S_1 \leftarrow S_0$ excitation energies for WT-Rh and the 30fold mutant from snapshots of a 520 ps QM1-MD simulation (OM2/MRCI, eV). Black lines are the excitation energy of the optimized structures.

for the enlarged BLA being a result of a hypsochromic shift and not vice versa. An enlarged BLA in the MD simulation has a higher excitation energy but much lower than expected with the slope in Figure 50.

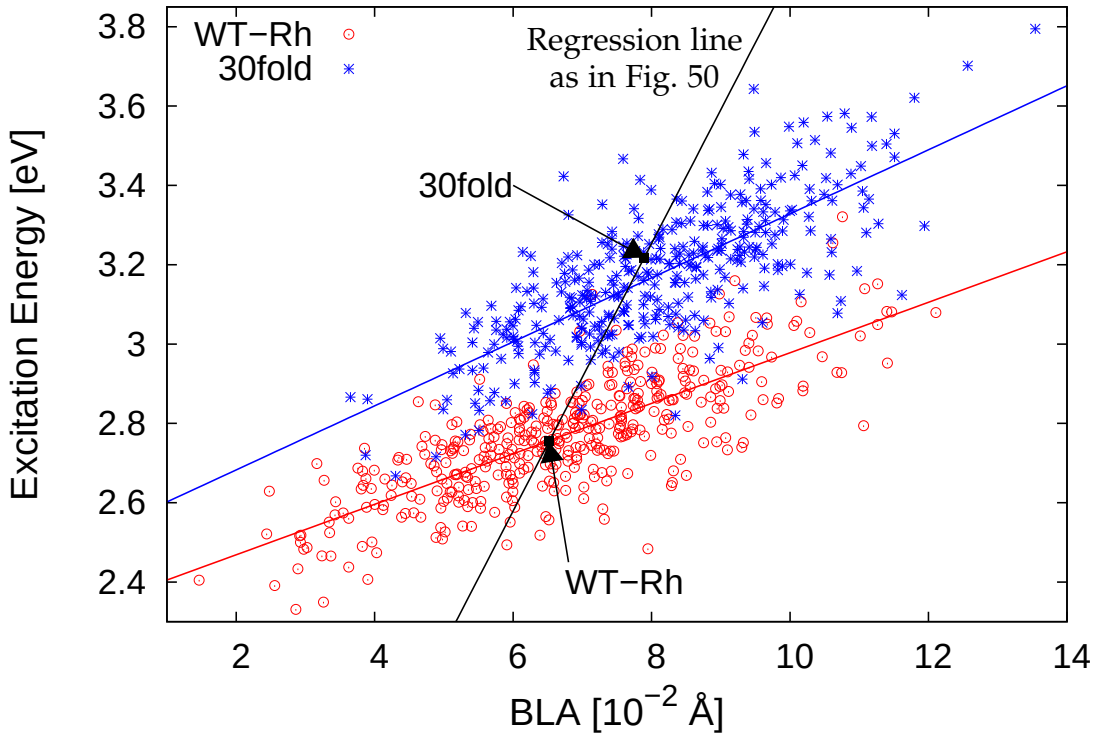


Figure 52: Correlation of excitation energies with the BLA of QM1-MD snapshots of WT-Rh and the 30fold mutant (OM2/MRCI, eV).

5.4.3 Effect of Protein Polarizability

To investigate the influence of the protein polarizability on the shifts in excitation energy, CHARMM point charges are substituted by an explicit polarization model.[49,50] The absolute excitation energies are lowered by 0.09 eV in average (see Table 58). The shifts relative to WT-Rh for the mutants are hardly affected (see Table 59).

For WT-Rh with QM4 and QM5_90 region the vertical excitation energy calculated with the polarized charges with respect to the S_0 charge distribution (ΔE^{vert}) is similar to the excitation energy with CHARMM charges (ΔE^{CHARMM}) (see Table 58). This small $\Delta\Delta E$ is a good evidence for the right choice of the QM region. For the medium sized QM2 region the difference is somewhat larger (0.05 eV). The polarization of Thr94 and Wat2b being in the MM region is largely different to the prepolarization of the CHARMM charges. For the 24fold and 30fold mutant the $\Delta\Delta E$ is slightly larger but this is probably due to the MM describtion of Wat45, Wat46 and the HBN above retinal. It is not manageble to increase the QM region this far.

The polarization shift ($\Delta E_{\text{pol}} - \Delta E_{\text{vert}}$) is similar for all mutants and all QM regions (-0.04 to -0.07 eV). Its origin is in the polarization of the charges with respect to the S_1 state and the energy win of the S_1 state due to this polarization. Thus, this shift is always a bathochromic one. It is slightly smaller for the 24fold and 30fold mutant and in part counterbalances the increase in $\Delta\Delta E$. Hence, the

Table 58: Absolute excitation energies $\Delta E_{S_1 \leftarrow S_0}$ with CHARMM charges and with polarization model, $\Delta\Delta E$, and polarization shift (SORCI(man), eV).

		$\Delta E_{S_1 \leftarrow S_0}$				
		CHARMM	vert	polar.h	$\Delta\Delta E$	pol-shift
	WT-Rh[a]	2.53	2.54	2.47	0.01	−0.07
QM5_90	10fold	2.78	2.76	2.70	−0.02	−0.06
	24fold	2.92	2.87	2.84	−0.05	−0.04
	30fold	2.98	2.94	2.89	−0.04	−0.05
	WT-Rh	2.53	2.54	2.47	0.01	−0.07
QM4	10fold	2.75	2.73	2.67	−0.02	−0.06
	24fold	2.89	2.86	2.81	−0.04	−0.04
	30fold	2.95	2.89	2.86	−0.05	−0.04
	WT-Rh	2.56	2.51	2.45	−0.05	−0.06
QM2	10fold	2.78	2.73	2.67	−0.05	−0.06
	24fold	2.91	2.85	2.80	−0.07	−0.04
	30fold	2.98	2.89	2.86	−0.09	−0.03

[a] For WT-Rh no QM5_90 calculations were performed because residue 90 is a glycine and the QM4 results are taken instead.

shifts in excitation energie with CHARMM charges and with the polarization model are very similar. Only for QM4 and QM5_90 the shifts with the polarization model are slightly smaller (0.02 – 0.03 eV, less than 10%).

Table 59: Shifts in excitation energy relative to WT-Rh with CHARMM charges and with polarization model (SORCI(man), eV)[a].

		CHARMM	polar.h
QM5_90	10fold	0.26	0.23
	24fold	0.39	0.37
	30fold	0.45	0.42
QM4	10fold	0.22	0.20
	24fold	0.36	0.35
	30fold	0.42	0.39
QM2	10fold	0.22	0.22
	24fold	0.35	0.35
	30fold	0.42	0.41

[a] QM5_90 shifts are relative to WT-Rh(QM4).

5.5 Summary and Conclusions

The quality of the structural model of the HB visual pigment can not be evaluated by only computational methods. If all orientations of mutated residues or unmutated residues close to mutations are right or if larger movements of helices occur when going from rhodopsin to HB can only be tested by X-ray crystallography of HB. In lack of such a structure, the resulting structure of this work is the best that can be found. The method of building the structure is very systematic and the final structure is reasonably stable in MD simulations.

The main task of this work is to produce a model of HB to study the color tuning mechanism between rhodopsin and HB. The remaining structural questions (e.g. the possible disulfide bridge between Cys87 and Cys299 (see page 112)) are not important for the excitation energy and the shift in excitation energy. There are some discrepancies in the calculated results in comparison to the experimental ones. These can come from errors in the built structures or from the measurements. The publication from Lin et al. giving the experimental results of some single and multiple mutants is from 1998 and shows some spectra that have a broad maximum in the absorption spectrum and some random noise. If the accuracy of these measurements is really 1 nm (0.005 eV) is not clear. The errors in the modeled structures can come from cooperative effects that were not found in the calculation or that are erroneously found.

The calculated shift in excitation energy for the 30fold mutant (0.42 eV) is sufficiently close to the experimental one for the HB pigment (0.5 eV) taking into

account that only 30 residues are mutated. A more exact model for the HB visual pigment would need a lot of additional work. Either a lot of manual work or homology modeling for the remaining mutations and long MD simulation with explicit water and lipids building the membrane. The model presented here is thought to give a good representation of the binding pocket of the HB pigment.

With this model and the other multiple mutant models, the decomposition of the color tuning shift is possible. The main part of the shift (81%) comes from the changed electrostatic effect of the binding pocket. The changed structure of the complex counterion is only responsible for 14% and finally the changed retinal conformation gives a shift of only 5% (see Table 54). This small number for the chromophore geometry is shown to be realistic (see Table 57). The strongly enlarged BLA is a result of the hypsochromic shift but not the reason for it.

It is shown that polarisation effects play a role for absolute excitation energies of rh and HB but the effect on the shift in excitation energy is very limited because the effect is similar for Rh and HB.

Chapter 6

Bibliography

(1) Weber, W.; Thiel, W. *Theor. Chem. Acc.* **2000**, *103*, 495–506.

(2) Neese, F. *ORCA - An ab initio, density functional and semiempirical program package*; Version 2.4 - Revision 41; Max Planck Institut fuer Strahlenchemie: Muelheim, Germany, 2005.

(3) Neese, F. *J. Chem. Phys.* **2003**, *119*, 9428–9443.

(4) Ferreira, K. N.; Iverson, T. M.; Maghlaoui, K.; Barber, J.; Iwata, S. *Science* **2004**, *303*, 1831–1838.

(5) Birge, R. R. *Ann. Rev. of Phys. Chem.* **1990**, *41*, 683–733.

(6) Hecht, S.; Shlaer, S.; Pirenne, M. *JGP* **1942**, *25*, 819–840.

(7) Young, R. *Sci. Am.* **1970**, *223*, 81–91.

(8) Zhukovsky, E. A.; Oprian, D. D. *Science* **1989**, *246*, 928–930.

(9) Sakmar, T. P.; Franke, R. R.; Khorana, H. G. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 8309–8313.

(10) Ridge, K. D.; Palczewski, K. *J. Biol. Chem.* **2007**, *282*, 9297–9301.

(11) Yan, E. C. Y.; Kazmi, M. A.; Ganim, Z.; Hou, J.-M.; Pan, D.; Chang, B. S. W.; Sakmar, T. P.; Mathies, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 9262–9267.

(12) Lüdeke, S.; Beck, M.; Yan, E. C. Y.; Sakmar, T. P.; Siebert, F.; Vogel, R. *J. Mol. Biol.* **2005**, *353*, 345–356.

(13) Kosower, E. M. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 1076–1080.

(14) Palczewski, K.; Kumasaka, T.; Hori, T.; Behnke, C. A.; Motoshima, H.; Fox, B. A.; Le Trong, I.; Teller, D. C.; Okada, T.; Stenkamp, R. E.; Yamamoto, M.; Miyano, M. *Science* **2000**, *289*, 739–745.

(15) Teller, D. C.; Okada, T.; Behnke, C. A.; Palczewski, K.; Stenkamp, R. E. *Biochemistry* **2001**, *40*, 7761–7772.

(16) Okada, T.; Fujiyoshi, Y.; Silow, M.; Navarro, J.; Landau, E. M.; Shichida, Y. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 5982–5987.

(17) Stenkamp, R. E.; Filipek, S.; Driessen, C. A. G. G.; Teller, D. C.; Palczewski, K. *Biochimica et Biophysica Acta - Biomembranes* **2002**, *1565*, 168–182.

(18) Kandori, H.; Shichida, Y.; Yoshizawa, T. *Biochemistry (Moscow)* **2001**, *66*, 1197–1209.

(19) Oprian, D. D. *Biochemistry* **1991**, *30*, 11367–11372.

(20) Fasick, J. I.; Lee, N.; Oprian, D. D. *Biochemistry* **1999**, *38*, 11593–11596.

(21) Wang, Z.; Asenjo, A. B.; Oprian, D. D. *Biochemistry* **1993**, *32*, 2125–2130.

(22) Berg, J. M.; Tymocsko, J. L.; Stryer, L. *Biochemistry*, 5th ed.; W.H. Freeman and Company: New York, 2002.

(23) Barth, A. *Progress In Biophysics & Molecular Biology* **2000**, *74*, 141–173.

(24) Seabra, G. D. M.; Walker, R. C.; Roitberg, A. E. *J. Phys. Chem. A* **2009**, *113*, 11938–11948.

(25) Tajkhorshid, E.; Baudry, J.; Schulten, K.; Suhai, S. *Biophys. J.* **2000**, *78*, 683–693.

(26) Vreven, T.; Morokuma, K. *J. Chem. Phys.* **2000**, *113*, 2969–2975.

(27) König, P.; Hoffmann, M.; Frauenheim, T.; Cui, Q. *J. Phys. Chem. B* **2005**, *109*, 9082–9095.

(28) Jensen, F. *Introduction to computational chemistry*; John Wiley & Sons, 2007.

(29) Szabó, A.; Ostlund, N. *Modern quantum chemistry: introduction to advanced electronic structure theory*; Dover Publications, 1996.

(30) Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136*, 864–871.

(31) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(32) Adamo, C.; Barone, V. *J. Chem. Phys.* **1999**, *110*, 6158–6170.

(33) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comp. Chem.* **1983**, *4*, 187–217.

(34) MacKerell, A. et al. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(35) Dinner, A. R.; Lopez, X.; Karplus, M. *Theor. Chem. Acc.* **2003**, *109*, 118–124.

(36) Simonson, T.; Archontis, G.; Karplus, M. *J. Phys. Chem. B* **1997**, *101*, 8349–8362.

(37) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B* **1998**, *58*, 7260–7268.

(38) Gaus, M.; Cui, Q.; Elstner, M. *J. Chem. Theo. Comp.* **2011**, *7*, 931–948.

(39) Yang, Y.; Yu, H.; York, D.; Cui, Q.; Elstner, M. *J. Phys. Chem. A* **2007**, *111*, 10861–10873.

(40) Cramer, C. *Essentials of computational chemistry: theories and models*; Wiley, 2004.

(41) Thiel, W. MNDO99, version 6.1; Max-Planck-Institut für Kohlenforschung: Mülheim an der Ruhr, Germany, 2006.

(42) Brooks, B. R.; Schaefer III, H. F. *The Journal of chemical physics* **1979**, *70*, 5092–5106.

(43) Hoffmann, M.; Wanko, M.; Strodel, P.; König, P. H.; Frauenheim, T.; Schulten, K.; Thiel, W.; Tajkhorshid, E.; Elstner, M. *J. Am. Chem. Soc.* **2006**, *128*, 10808–10818.

(44) Wanko, M.; Hoffmann, M.; Strodel, P.; Koslowski, A.; Thiel, W.; Neese, F.; Frauenheim, T.; Elstner, M. *J. Phys. Chem. B* **2005**, *109*, 3606–3615.

(45) Mogensen, B. J.; Rettrup, S. *Int. J. Quantum Chem.* **1992**, *44*, 1045–1056.

(46) Miralles, J.; Castell, O.; Caballol, R.; Malrieu, J.-P. *Chem. Phys.* **1993**, *172*, 33 – 43.

(47) Schäfer, A.; Horn, H.; Ahlrichs, R. *J. Chem. Phys.* **1992**, *97*, 2571–2577.

(48) Altun, A.; Yokoyama, S.; Morokuma, K. *J. Phys. Chem. B* **2008**, *112*, 16883–16890.

(49) Wanko, M.; Hoffmann, M.; Frauenheim, T.; Elstner, M. *J. Phys. Chem. B* **2008**, *112*, 11462–11467.

(50) Wanko, M.; Hoffmann, M.; Frähmcke, J.; Frauenheim, T.; Elstner, M. *J. Phys. Chem. B* **2008**, *112*, 11468–11478.

(51) Sekharan, S.; Sugihara, M.; Buss, V. *Angew. Chem., Int. Ed.* **2007**, *46*, 269–271.

(52) Okada, T.; Sugihara, M.; Bondar, A.-N.; Elstner, M.; Entel, P.; Buss, V. *J. Mol. Biol.* **2004**, *342*, 571–583.

(53) Li, J.; Edwards, P. C.; Burghammer, M.; Villa, C.; Schertler, G. F. *J. Mol. Biol.* **2004**, *343*, 1409–1438.

(54) Standfuss, J.; Xie, G.; Edwards, P. C.; Burghammer, M.; Oprian, D. D.; Schertler, G. F. *J. Mol. Biol.* **2007**, *372*, 1179–1188.

(55) Ruprecht, J. J.; Mielke, T.; Vogel, R.; Villa, C.; Schertler, G. F. X. *EMBO Journal* **2004**, *23*, 3609–3620.

(56) Terakita, A.; Yamashita, T.; Shichida, Y. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 14263–14267.

(57) Yan, E. C. Y.; Kazmi, M. A.; De, S.; Chang, B. S. W.; Seibert, C.; Marin, E. P.; Mathies, R. A.; Sakmar, T. P. *Biochemistry* **2002**, *41*, 3620–3627.

(58) Frähmcke, J. S.; Wanko, M.; Phatak, P.; Mroginski, M. A.; Elstner, M. *J. Phys. Chem. B* **2010**, *114*, 11338–11352.

(59) Kakitani, H.; Kakitani, T.; Rodman, H.; Honig, B. *Photochem. Photobiol.* **1985**, *41*, 471–479.

(60) Honig, B.; Dinur, U.; Nakanishi, K.; Balogh-Nair, V.; Gawinowicz, M. A.; Arnaboldi, M.; Motto, M. C. *J. Am. Chem. Soc.* **1979**, *101*, 7084–7086.

(61) Shriver, J.; Mateescu, G.; Fager, R.; Torchia, D.; Abrahamson, E. W. *Nature* **1977**, *270*, 271–273.

(62) Birge, R. R.; Murray, L. P.; Pierce, B. M. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 4117–4121.

(63) Nielsen, I. B.; Lammich, L.; Andersen, L. H. *Phys. Re. Lett.* **2006**, *96*, 018304.

(64) Ebrey, T. G.; Honig, B. *Proc. Natl. Acad. Sci. U.S.A.* **1972**, *69*, 1897–1899.

(65) Tomasello, G.; Olaso-Gonzales, G.; Altoe, P.; Stenta, M.; Serrano-Andres, L.; Merchan, M.; Orlandi, G.; Bottoni, A.; Garavelli, M. *J. Am. Chem. Soc.* **2009**, *131*, 5172–5186.

(66) Fahmy, K.; Jager, F.; Beck, M.; Zvyaga, T. A.; Sakmar, T. P.; Siebert, F. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 10206–10210.

(67) Nagata, T.; Terakita, A.; Kandori, H.; Shichida, Y.; Maeda, A. *Biochemistry* **1998**, *37*, 17216–17222.

(68) Nathans, J. *Biochemistry* **1990**, *29*, 937–942.

(69) Lewis, J. W.; Szundi, I.; Kazmi, M. A.; Sakmar, T. P.; Kliger, D. S. *Biochemistry* **2004**, *43*, 12614–12621.

(70) Janz, J. M.; Farrens, D. L. *J. Biol. Chem.* **2004**, *279*, 55886–55894.

(71) Ahuja, S.; Smith, S. *Trends Pharmacol. Sci.* **2009**, *30*, 494–502.

(72) Altun, A.; Yokoyama, S.; Morokuma, K. *J. Phys. Chem. B* **2008**, *112*, 6814–6827.

(73) Andruniow, T.; Ferre, N.; Olivucci, M. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 17908–17913.

(74) Buss, V.; Sugihara, M.; Entel, P.; Hafner, J. *Angew. Chem., Int. Ed.* **2003**, *42*, 3245–3247.

(75) Coto, P. B.; Strambi, A.; Ferre, N.; Olivucci, M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 17154–17159.

(76) Ferre, N.; Olivucci, M. *J. Am. Chem. Soc.* **2003**, *125*, 6868–6869.

(77) Fujimoto, K.; Hasegawa, J. .; Hayashi, S.; Kato, S.; Nakatsuji, H. *Chem. Phys. Lett.* **2005**, *414*, 239–242.

(78) Fujimoto, K.; Hasegawa, J. .; Hayashi, S.; Nakatsuji, H. *Chem. Phys. Lett.* **2006**, *432*, 252–256.

(79) Sugihara, M.; Buss, V.; Entel, P.; Elstner, M.; Frauenheimn, T. *Biochemistry* **2002**, *41*, 15259–15266.

(80) Sugihara, M.; Buss, V.; Entel, P.; Hafner, J. *J. Phys. Chem. B* **2004**, *108*, 3673–3680.

(81) Sugihara, M.; Hufen, J.; Buss, V. *Biochemistry* **2006**, *45*, 801–810.

(82) Sekharan, S.; Sugihara, M.; Weingart, O.; Okada, T.; Buss, V. *J. Am. Chem. Soc.* **2007**, *129*, 1052–1054.

(83) Virshup, A. M.; Punwong, C.; Pogorelov, T. V.; Lindquist, B. A.; Ko, C.; Martinez, T. J. *J. Phys. Chem. B* **2009**, *113*, 3280–3291.

(84) Röhrig, U. F.; Guidoni, L.; Rothlisberger, U. *Biochemistry* **2002**, *41*, 10799–10809.

(85) Schreiber, M.; Buss, V.; Sugihara, M. *J. Chem. Phys.* **2003**, *119*, 12045–12048.

(86) Martinez-Mayorga, K.; Pitman, M. C.; Grossfield, A.; Feller, S. E.; Brown, M. F. *J. Am. Chem. Soc.* **2006**, *128*, 16502–16503.

(87) Grossfield, A.; Pitman, M. C.; Feller, S. E.; Soubias, O.; Gawrisch, K. *J. Mol. Biol.* **2008**, *381*, 478–486.

(88) Hall, K. F.; Vreven, T.; Frisch, M. J.; Bearpark, M. J. *J. Mol. Biol.* **2008**, *383*, 106–121.

(89) Röhrig, U. F.; Sebastiani, D. *J. Phys. Chem. B* **2008**, *112*, 1267–1274.

(90) Sekharan, S. *Photochem. Photobiol.* **2009**, *85*, 517–520.

(91) Sekharan, S.; Buss, V. *J. Am. Chem. Soc.* **2008**, *130*, 17220–17221.

(92) Mathies, R.; Stryer, L. *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73*, 2169–2173.

(93) Wada, M.; Sakurai, M.; Inoue, Y.; Chûjô, R. *Magn. Reson. Chem.* **1992**, *30*, 831–836.

(94) Han, M.; DeDecker, B. S.; Smith, S. O. *Biophys. J.* **1993**, *65*, 899–906.

(95) Han, M.; Smith, S. O. *Biophys. Chem.* **1995**, *56*, 23–29.

(96) Verhoeven, M. A.; Creemers, A. F. L.; Bovee-Geurts, P. H. M.; De Grip, W. J.; Lugtenburg, J.; de Groot, H. J. M. *Biochemistry* **2001**, *40*, 3282–3288.

(97) Creemers, A. F. L.; Kiihne, S.; Bovee-Geurts, P. H. M.; DeGrip, W. J.; Lugtenburg, J.; de Groot, H. J. M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 9101–9106.

(98) Menon, S. T.; Han, M.; Sakmar, T. P. *Physiol. Rev.* **2001**, *81*, 1659.

(99) Salgado, G. F. J.; Struts, A. V.; Tanaka, K.; Krane, S.; Nakanishi, K.; Brown, M. F. *J. Am. Chem. Soc.* **2006**, *128*, 11067–11071.

(100) Eilers, M.; Reeves, P. J.; Ying, W.; Khorana, H. G.; Smith, S. O. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 487–492.

(101) Page, C. S.; Olivucci, M. *J. Comput. Chem.* **2003**, *24*, 298–309.

(102) Keal, T.; Wanko, M.; Thiel, W. *Theor. Chem. Acc.* **2009**, *123*, 145–156.

(103) TURBOMOLE V5-10, 14 Jan 2008, Copyright (C) 2008 University of Karlsruhe.

(104) ChemShell, a Computational Chemistry Shell, see www.chemshell.org.

(105) Sherwood, P. et al. *J. Mol. Struct. (Theochem)* **2003**, *632*, 1–28.

(106) Magdo, I.; Nemeth, K.; Mark, F.; Hildebrandt, P.; Schaffner, K. *J. Phys. Chem. A* **1999**, *103*, 289–303.

(107) Mroginski, M. A.; Murgida, D. H.; Hildebrandt, P. *J. Phys. Chem. A* **2006**, *110*, 10564–10574.

(108) Mroginski, M. A.; Mark, F.; Thiel, W.; Hildebrandt, P. *Biophys. J.* **2007**, *93*, 1885–1894.

(109) Adamo, C.; Barone, V. *Chem. Phys. Lett.* **1998**, *298*, 113–119.

(110) Ditchfield, R. *Mol. Phys.* **1974**, *27*, 789–807.

(111) Frisch, M. J. et al. *Gaussian 03, Revision B.04*.

(112) Tobias, D. J.; Brooks III, C. L. *J. Chem. Phys.* **1988**, *89*, 5115–5127.

(113) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420–1426.

(114) Tobias, D. J.; Brooks III, C. L. *Chem. Phys. Lett.* **1987**, *142*, 472–476.

(115) Tobias, D. J.; Brooks III, C. L. *J. Phys. Chem.* **1992**, *96*, 3864–3870.

(116) Verlet, L. *Phys. Rev.* **1967**, *159*, 98–103.

(117) Lin, S. W.; Kochendoerfer, G. G.; Carroll, K. S.; Wang, D.; Mathies, R. A.; Sakmar, T. P. *J. Biol. Chem.* **1998**, *273*, 24583–24591.

(118) Kochendoerfer, G. G.; Wang, Z.; Oprian, D. D.; Mathies, R. A. *Biochemistry* **1997**, *36*, 6577–6587.

(119) Carravetta, M.; Zhao, X.; Johannessen, O. G.; Lai, W. C.; Verhoeven, M. A.; Bovee-Geurts, P. H. M.; Verdegem, P. J. E.; Kiihne, S.; Luthman, H.; de Groot, H. J. M.; deGrip, W. J.; Lugtenburg, J.; Levitt, M. H. *J. Am. Chem. Soc.* **2004**, *126*, 3948–3953.

(120) Phatak, P.; Ghosh, N.; Yu, H.; Cui, Q.; Elstner, M. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 19672–19677.

(121) Phatak, P.; Frähmcke, J. S.; Wanko, M.; Hoffmann, M.; Strodel, P.; Smith, J. C.; Suhái, S.; Bondar, A. N.; Elstner, M. *J. Am. Chem. Soc.* **2009**, *131*, 7064–7078.

(122) Sander, W.; Gantenberg, M. *Spectrochim. Acta - Part A* **2005**, *62*, 902–909.

(123) Mroginski, M. A.; Kaminski, S.; Hildebrandt, P. *ChemPhysChem* **2010**, *11*, 1265–1274.

(124) Nielsen, I. B.; Lammich, L.; Andersen, L. H. *Phys. Rev. Lett.* **2006**, *96*, 018304.

(125) Yamazaki, M.; Goodisman, J.; Birge, R. R. *J. Chem. Phys.* **1998**, *108*, 5876–5887.

(126) Chizhov, I.; Schmies, G.; Seidel, R.; Sydor, J. R.; Lüttenberg, B.; Engelhard, M. *Biophys. J.* **1998**, *75*, 999–1009.

(127) Wanko, M. *Dissertation*, TU Braunschweig, 2009.

(128) Stuart, J. A.; Vought, B. W.; Zhang, C.-F.; Birge, R. R. *Biospectroscopy* **1995**, *1*, 9–28.

(129) Birge, R. R.; Knox, B. E. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 9105–9107.

(130) Harbison, G. S.; Mulder, P. P. J.; Pardoen, J. A. *J. Am. Chem. Soc.* **1985**, *107*, 4809–4816.

(131) Smith, S. O.; Palings, I.; Miley, M. E.; Courtin, J.; de Groot, H.; Lugtenburg, J.; Mathies, R. A.; Griffin, R. G. *Biochemistry* **1990**, *29*, 8158–8164.

(132) Smith, S. O.; Courtin, J.; de Groot, H.; Gebhard, R.; Lugtenburg, J. *Biochemistry* **1991**, *30*, 7409–7415.

(133) Mollevanger, L. C. P. J.; Kentgens, A. P. M.; Pardoen, J. A.; Courtin, J. M. L.; Veeman, W. S.; Lugtenburg, J.; Degrip, W. J. *Eur. J. Biochem.* **1987**, *163*, 9–14.

(134) Spooner, P. J. R.; Sharples, J. M.; Verhoeven, M. A.; Lugtenburg, J.; Glaubitz, C.; Watts, A. *Biochemistry* **2002**, *41*, 7549–7555.

(135) Smith, S. O.; de Groot, H. J. M.; Gebhard, R.; Courtin, J. M. L.; Lugtenburg, J.; Herzfeld, J.; Griffin, R. G. *Biochemistry* **1989**, *28*, 8897–8904.

(136) Concistre, M.; Gansmuller, A.; McLean, N.; Johannessen, O. G.; Montesinos, I. M.; Bovee-Geurts, P. H. M.; Verdegem, P.; Lugtenburg, J.; Brown, R. C. D.; DeGrip, W. J.; Levitt, M. H. *J. Am. Chem. Soc.* **2008**, *130*, 10490–10491.

(137) Gascon, J. A.; Sproviero, E. M.; Batista, V. S. *J. Chem. Theory Comput.* **2005**, *1*, 674–685.

(138) Sandberg, M. N.; Amora, T. L.; Ramos, L. S.; Chen, M.-H.; Knox, B. E.; Birge, R. R. *J. Am. Chem. Soc.* **2011**, *133*, 2808–2811.

(139) Frähmcke, J. S.; Wanko, M.; Elstner, M. *J. Phys. Chem. B* **2012**, *116*, 3313–3321.

(140) Zöllner, M. J.; Frähmcke, J. S.; Elstner, M.; Jahn, U.; Jones, P. G.; Becker, E.; Kowalsky, W.; Johannes, H.-H. *Macrom. Chem. and Phys.* **2010**, *211*, 359–371.

(141) Holleboom, C. P. *Diploma Thesis*, TU Braunschweig, 2010.

Appendix A

Details on OM2 Calculations

Size of the Active MO window

The size of this active window in OM2/MRCI calculations was benchmarked to have all π orbitals that are important for the excited states in the active window. Asymmetric windows were excluded to make it simpler.

Table A-1: Variation of the active window in an OM2/MRCI calculation (QM4). Effect on $S_1 \leftarrow S_0$ excitation energy (eV) and calculation time (h).

(a) WT-Rh

	$\Delta E_{S_1 \leftarrow S_0}$	time
12–12	2.84	0:18
13–13	2.83	0:31
14–14	2.82	0:49
15–15	2.82	1:14
16–16	2.82	1:53
17–17	2.81	2:44
18–18	2.81	3:55

(b) 30fold mutant

	$\Delta E_{S_1 \leftarrow S_0}$	time
12–12	3.23	0:18
13–13	3.22	0:33
14–14	3.22	0:52
15–15	3.22	1:20
16–16	3.21	2:16
17–17	3.21	3:18
18–18	3.20	4:42

For WT-Rh the best choice seems to be an active window of 28 MOs (14–14) but for the 30fold mutant and other structures small discrepancies were observed and thus a larger window of 32 MOs (16–16, `ICI1=16 ICI2=16`) was used to have consistent results without the need to check all calculations for convergence. A calculation time of less than three hours is still acceptable.

Number of Starting References

The number of starting references is chosen large (five) to have a low probability that the CISD calculation has to be done again and to have a comparable

number of references for all calculations. Sometimes one or two more references are chosen. To have this very seldomly, a low value for the threshold is used (`CISELR=9000`). Only if the sum of weighting factors for one of the calculated states is smaller than 90%, more references are added. For an example calculation the result is shown in Table A-2

Table A-2: Main configurations and their weight in an OM2/MRCI calculation (30fold mutant, QM4)

(a) WT-Rh

S_0		S_1	
2200	91.5%	2110	83.3%
2020	1.1%	1201	3.3%
		2020	1.8%
		1111	1.6%

(b) 30fold mutant

S_0		S_1	
2200	91.6%	2110	83.6%
2020	1.1%	1201	4.6%
		2020	1.1%
		1111	1.0%

The threshold for adding more and more references can be increased to guarantee a balanced inclusion of correlation effects for both states. But with many references the calculation time gets much too long and the amount of memory needed increases drastically. Increasing the threshold to 91, 92, 93, and 94% gives calculation more than two-times longer due to more references in the second MR-CISD calculation.

Table A-3: Effect of increased threshold for selecting more references in an OM2/MRCI calculation on the number of references N, the $S_1 \leftarrow S_0$ excitation energy (eV) and the calculation time (WT-Rh, QM4, h).

`CISELR`	N	new references	$\Delta E_{S_1 \leftarrow S_0}$	time
9000	5	–	2.82	1:53
9100	6	2221221 1010000	2.8	7:23
9200	8	2221222 0100000, 2222212 1000000	2.81	7:59
9300	11	2221222 0010000, 2222220 1100000, 2221212 2000000	2.81	10:17
9400	16	0222221 1000002, 2222201 1020000, 2222211 1010000, 2222220 0200000, 2221221 1000100	–	–[a]

[a] Calculation ended with an error after ca. 2 hours.

Comparison to SORCI

Results with the QM4 region and with all external charges, the difference between OM2/MRCI and SORCI is about ≈0.29 eV for WT-Rh and similar mutants and ≈0.26 eV for blue mutants as the 30fold mutant. Thus small shifts can be predicted very nicely with OM2. For very blue shifted structures the shifts calculated with the more accurate SORCI method are slightly larger.

Table A-4: OM2/MRCI and SORCI results ($\Delta E_{S_1 \leftarrow S_0}$, eV) of some WT-Rh and mutant structures calculated with different QM regions and electrostatics.

structure	QM	charges	OM2/MRCI	SORCI	Δ
WT-Rh	QM1	w/o	2.32	1.98	0.34
WT-Rh	QM1	with	2.75	2.53	0.22
WT-Rh	QM2	w/o	3.00	2.79	0.21
WT-Rh	QM2	with	2.83	2.56	0.27
WT-Rh	QM4	w/o	2.95	2.76	0.19
WT-Rh	QM4	with	2.82	2.53	0.29
A292S	QM4	with	2.86	2.57	0.29
A117G	QM4	with	2.81	2.52	0.29
h7	QM4	with	2.86	2.56	0.30
h36	QM4	with	2.96	2.68	0.28
10fold	QM4	with	3.03	2.75	0.28
24fold	QM4	with	3.15	2.89	0.26
28fold	QM4	with	3.15	2.87	0.27
30fold	QM4	with	3.21	2.95	0.26

Appendix B

Details on SORCI Calculations (Selection of π Orbitals)

As decribed in the methods part the selection of the π orbitals for the SORCI calculation is a crucial point. It is needed, because the restricted active space (RAS) is used first to build the reference space and afterwards to determine all configurations being included in the DDCI2 or DDCI3 space. The twelve frontier orbitals are of special interest and are thus privileged. It is important that these twelve orbitals are the π orbitals of retinal and not something different, e.g., σ orbitals or π orbitals on other parts of the QM region.

After a normal SCF calculation and the improvement of the virtual orbitals the order of MOs is strictly by energy. But only some orbitals are pure π orbitals. Most of them are mixtures of π and σ orbitals or parts of the MOs are located away from retinal (on the counterion or other residues). If a SORCI calculation would be done with these orbitals, the π orbitals outside these twelve frontier orbitals are less privileged as they should be and thus the excitation energies are wrong and even the states can be qualitatively wrong. Thus the MOs have to be reordered to have twelve pure π orbitals in the RAS.

Three different methods are described in the following: Manual reordering and automatic building of natural MOs with two methods called nat3 and nat19.

B.1 Manual Reordering (man)

Reordering can be interchanging two orbital numbers to put an unwanted orbital out and a wanted orbital into the twelve orbital frontier region. Orbitals being a mixture of a π and a different orbital have to be converted to a pure π and a pure different orbital. This is done by building a linear combination of both orbitals resulting in two new orbitals being the sum and difference of the old ones. In ORCA this mixing is implemented as a kind of rotation. Mathematically

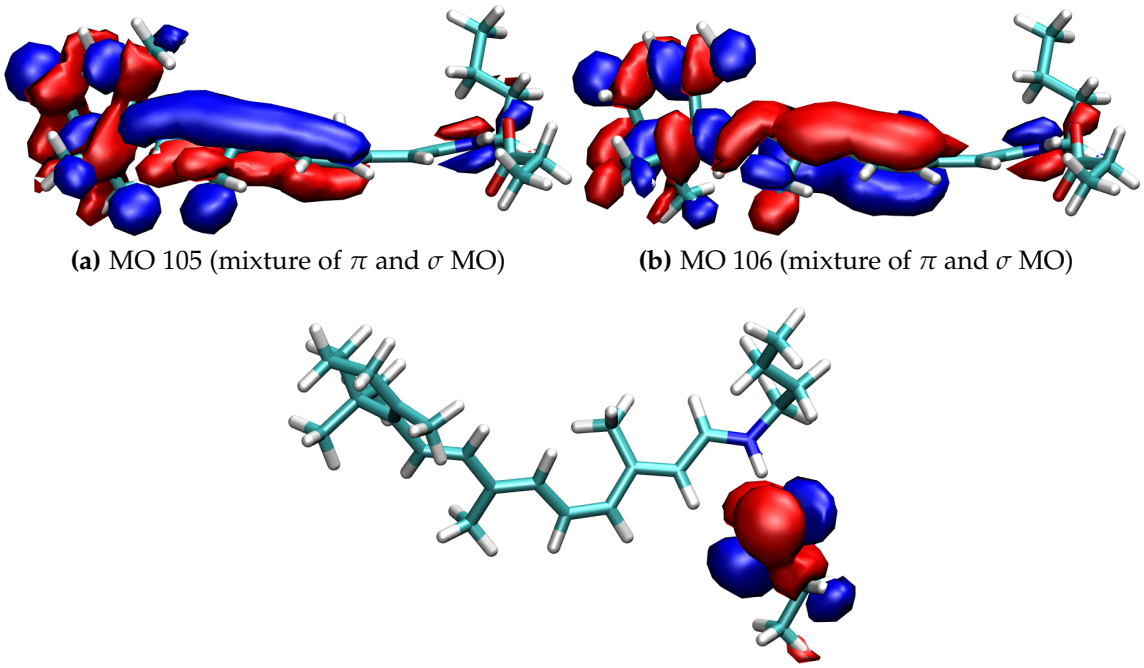
the relation between the linear combination and rotation can be explained by

$$\Psi_1' = c_1 \cdot \Psi_1 - c_2 \cdot \Psi_2 \tag{39}$$
$$\Psi_2' = c_2 \cdot \Psi_1 + c_1 \cdot \Psi_2 \tag{40}$$

with $c_1 = \cos\phi$ and $c_2 = \sin\phi$. A rotation by $\phi = 90°$ is an exchange. $\phi = 45°$ is a 1:1 ($\frac{1}{2}\sqrt{2} : \frac{1}{2}\sqrt{2}$) mixture of both orbitals. With other values for ϕ other mixtures can be built. These reordering has an effect on the orbital energies only if orbitals are mixed. After reordering, the twelve frontier orbitals can have a lower energy than other occupied MOs or higher energies than virtual orbitals, but this has no effect on the calculation and is physically correct.

Example for mixing and exchange

Figure A-1 shows some MOs of the 30fold mutant with the QM2 region after the SCF calculation. The HOMO is MO 113. MOs 108 to 119 are included in the RAS and should be pure π MOs. MOS 105 and 106 are mixtures of a π and a σ MO and MO 109 is a pure π MO but on the counterion. The task is to mix MOS 105 and 106 and exchange the sum of both with MO 109 to have a pure π MO at MO 109. This is done with two rotations `{105,106,45}` and `{105,109,90}` in the SORCI input. The MOs after this rotation are shown in Figure A-2.



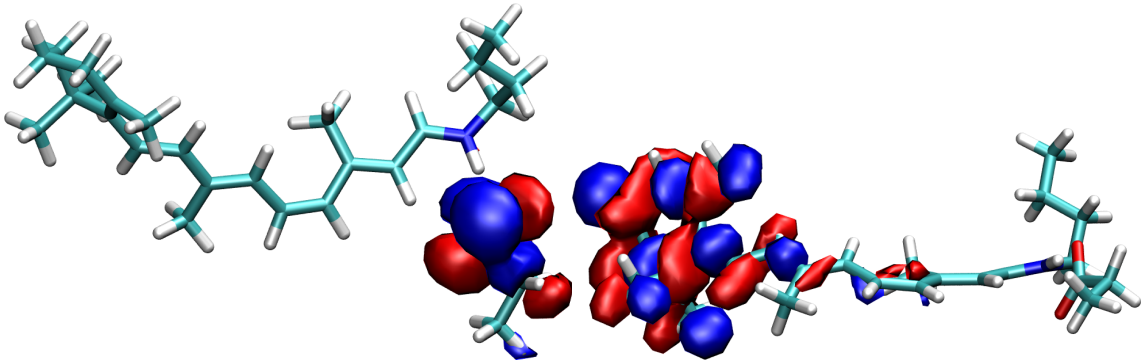
(a) MO 105 (mixture of π and σ MO)

(b) MO 106 (mixture of π and σ MO)

(c) MO 109 (π MO on the counterion)

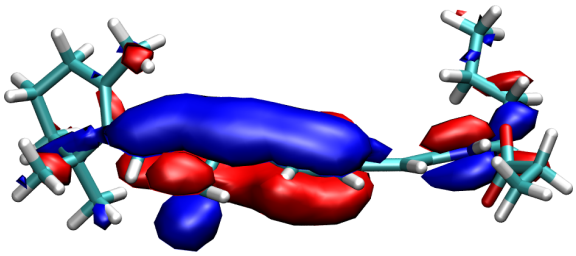
Figure A-1: MOs after SCF, before rotation (30fold mutant, QM2).

Figure A-3 and A-4 show the twelve frontier MOs after all rotations.



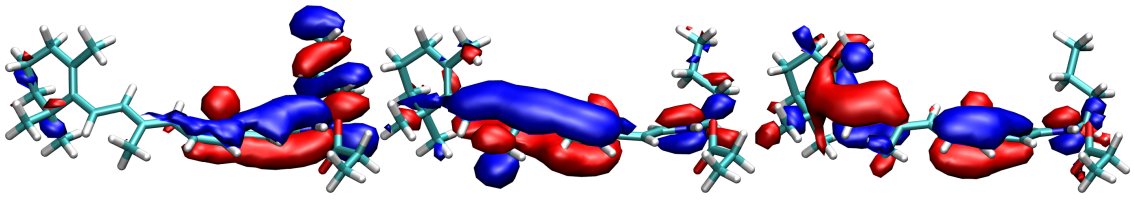
(a) MO 105 (π MO on the counterion)

(b) MO 106 (pure σ MO)

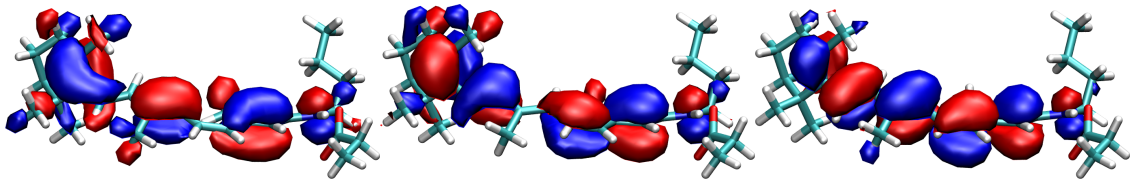


(c) MO 109 (pure π MO)

Figure A-2: MOs after rotation (30fold mutant, QM2).

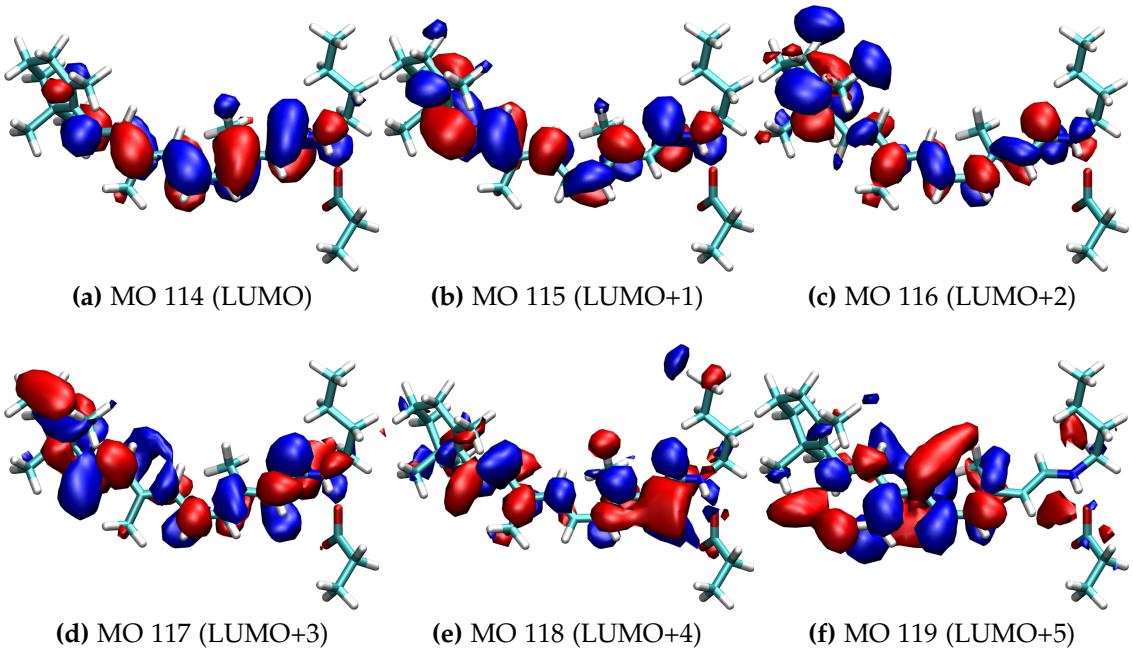


(a) MO 108 (HOMO–5) **(b)** MO 109 (HOMO–4) **(c)** MO 110 (HOMO–3)



(d) MO 111 (HOMO–2) **(e)** MO 112 (HOMO–1) **(f)** MO 113 (HOMO)

Figure A-3: Six occupied frontier MOs after rotation (30fold mutant, QM2).



(a) MO 114 (LUMO) **(b)** MO 115 (LUMO+1) **(c)** MO 116 (LUMO+2)

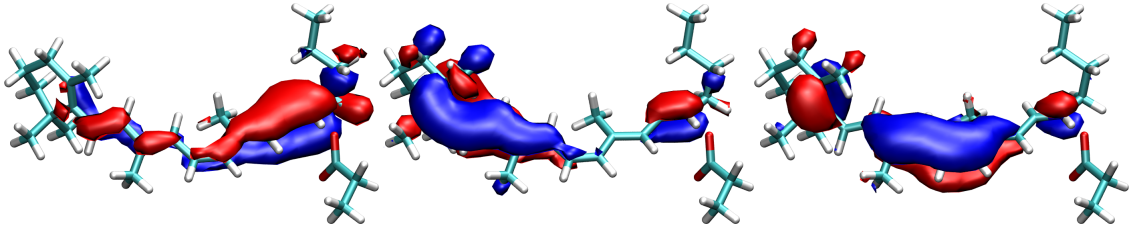
(d) MO 117 (LUMO+3) **(e)** MO 118 (LUMO+4) **(f)** MO 119 (LUMO+5)

Figure A-4: Six virtual frontier MOs after rotation (30fold mutant, QM2).

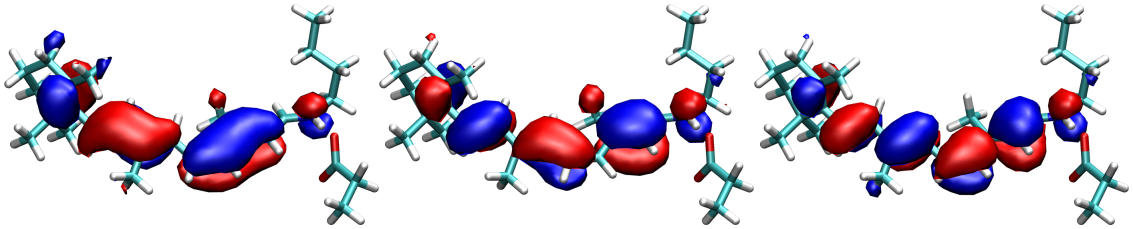
B.2 Automatic Building of Natural MOs (nat3)

The input orbitals for the SORCI calculataion do not have to be SCF MOs or linear combination of those. Also natural MOs from a preliminary calculation can be used as input. The nat3 methods builds natural orbitals with a short MRDDCI2 calculation. In contrast to a normal SORCI calculation a small CAS(2,2) (instead of RAS(12: 5 2 /2/ 5 2)) is used as starting point for building the reference space and all excited configuration based on that. This CAS consists of three configurations: 22222(20)00000, 22222(11)00000, and 22222(02)00000. The HOMO and LUMO are the only orbitals included in this CAS and both are always π MOs after the SCF calculation. Therefore, no manual inspection or rotation of MOs is needed and the nat3 method can be used as an automatic SORCI method. To further accelarate the calculation, a smaller energy window is used: `ewin -0.6,0.25` (instead of `ewin -3.0,5.0`). The natural MOs produced by the MRDDCI2 calculations are read in in the SORCI calculation.

Figure A-5 and A-6 show the twelve frontier MOs after the MRDDCI2 calculation.

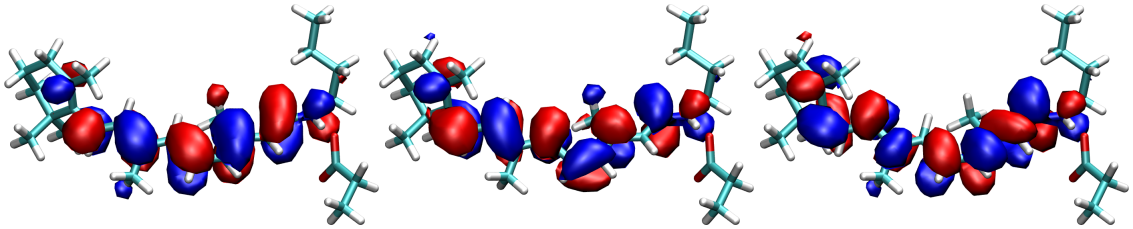


(a) MO 108 (HOMO–5) **(b)** MO 109 (HOMO–4) **(c)** MO 110 (HOMO–3)

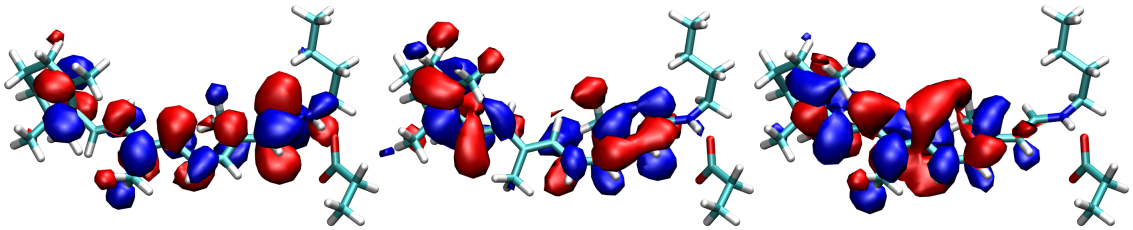


(d) MO 111 (HOMO–2) **(e)** MO 112 (HOMO–1) **(f)** MO 113 (HOMO)

Figure A-5: Six occupied frontier MOs after MRDDCI2 calculation (30fold mutant, QM2).



(a) MO 114 (LUMO) **(b)** MO 115 (LUMO+1) **(c)** MO 116 (LUMO+2)



(d) MO 117 (LUMO+3) **(e)** MO 118 (LUMO+4) **(f)** MO 119 (LUMO+5)

Figure A-6: Six virtual frontier MOs after MRDDCI2 calculation (30fold mutant, QM2).

B.3 Automatic Building of Natural MOs (nat19)

The second automatic method to build pure π orbitals as input for the SORCI calculation is very similar to the nat3 method. It takes a CAS(4,4) instead of a CAS(2,2). CAS(4,4) consists of 19 configurations.

Table A-5: Configurations in a CAS(4,4).

HF	2222(2200)0000			
single	2222(2110)0000	2222(2101)0000	2222(1210)0000	2222(1201)0000
double	2222(2020)0000	2222(2011)0000	2222(2002)0000	2222(1120)0000
	2222(1111)0000	2222(1102)0000	2222(0220)0000	2222(0211)0000
	2222(0202)0000			
triple	2222(1021)0000	2222(1012)0000	2222(0121)0000	2222(0112)0000
quadruple	2222(0022)0000			

Additionally to the larger CAS, the MRDDCI3 method is used to give slightly better natural orbitals. The precalculation in nthe nat19 method only takes ten minutes longer than the calculation with the nat3 method. The reference space is more than 6 times larger but the most time-consuming step (diagonalization of the CI space) is only slightly slower because most configurations are not important and are thus not selected.

Usually HOMO–1 and LUMO+1 are pure π MOs without rotation. Thus a nat19 calculation can be used as an automatic SORCI method. Only for calculations with a large QM region consisting an aromatic residue (Trp265 or Tyr265), HOMO–1 has to be rotated to have a pure π MO.

The natural MOs are very similar to the ones of a nat3 calculation and are thus not shown.

B.4 Comparison of the Three Methods

In the beginning of the work, the automatic nat3 method was used. But the larger the shift to WT-Rh became the larger the discrepancy between SORCI calculations with manual selection of π orbitals and with the nat3 method became. Excitation energies with the nat3 method are generally 0.02 eV lower in energy but for small shifts in excitation energy the shifts are the same for both methods. But for all multiple mutants starting from h237 the shifts are also 0.02 to 0.05 eV lower than with manual selection.

At this point a new automatic method was needed and was found in the nat19 method giving excitation energies very close to the man-results (+0.005 eV) and shifts that are also very similar (+0.002 eV) (see Figure A-6).

Table A-6: Comparison of shifts in excitation energy calculated with SORCI (QM4) with manual selection (man), and the nat3 and nat19 method (eV).

	Shift to WT-Rh			Δ to man	
	man	nat3	nat19	nat3	nat19
M86L	0.00	-0.01	-0.01	0.00	-0.01
V87C	0.02	0.01	0.01	-0.01	-0.01
G90S	0.05	0.03	0.04	-0.02	-0.01
A117G	-0.02	-0.01	-0.01	0.01	0.01
E122L	-0.02	-0.02	-0.02	0.00	0.00
A124T	0.03	0.02	0.04	-0.01	0.02
W265Y	0.06	0.06	0.06	-0.01	0.00
A292S	0.04	0.04	0.04	0.00	0.00
A295S	-0.01	-0.01	-0.01	0.00	0.01
A299C	-0.01	-0.01	0.00	0.01	0.01
h2′	0.04	0.02	0.02	-0.02	-0.02
h2	0.03	0.02	0.03	-0.01	-0.01
h3	0.03	0.02	0.03	-0.01	0.01
h7	0.04	0.03	0.04	0.00	0.01
h27	0.05	0.05	0.06	0.00	0.01
h36	0.15	0.13	0.16	-0.02	0.01
h237	0.09	0.07	0.08	-0.02	-0.01
h267	0.15	0.13	0.16	-0.02	0.00
h267+117	0.17	0.14	0.18	-0.03	0.01
h267+122	0.15	0.14	0.16	-0.01	0.01
h267+124	0.16	0.15	0.17	-0.01	0.00
9fold	0.21	0.17	0.22	-0.04	0.02
10fold	0.22	0.17	0.22	-0.05	0.00
24fold	0.25	0.20	0.26	-0.05	0.01
28fold	0.25	0.22	0.26	-0.03	0.01
28fold+189	0.30	0.25	0.29	-0.04	0.00
28fold+125	0.30	0.26	0.29	-0.04	-0.01
30fold	0.33	0.29	0.31	-0.04	-0.03
		Mean Difference		-0.02	0.002

Appendix C

Details on SORCI Calculations (Number of States to Solve)

The calculation of excitation energies with the SORCI method is problematic for all cases where the charges of the counterion are not included. These are only calculations using the small QM region (QM1). In extended QM regions the counter ion is always included. Calculations without all external charges or without Glu113 or without Glu113, Wat2b and residue 94 give a good excitation energy for the $S_1 \leftarrow S_0$ excitation but the third state (S_2) calculated by SORCI is not the same as in all other cases. Additional calculations solving for more than 3 states give the „normal“ S_2 state as the fourth state with a reasonable energy. Table A-7 gives the main configurations building the calculated states for WT-Rh for the QM1 region without and with external charges.

The S_1 state is qualitatively the same with and without external charges and independent of the number of solved states. The $S_1 \leftarrow S_0$ excitation energy is hardly changed. For higher states there are qualitative changes and also the excitation energies are changed. When looking at higher states the number of states to be solved in the SORCI calculation, but for the $S_1 \leftarrow S_0$ excitation energy a standard calculation with three states is sufficient.

Table A-8 shows the data for the QM4 region. Qualitative changes in the states when solving for more states are smaller. Energetical changes are observable but the shifts in excitation energy between WT-Rh and the 30fold mutant are nearly unchanged.

As a conclusion, care has to be take when using the SORCI method if higher excited states are of interest. Esspecially for calculations without the counterion the calculated S_2 state can be different to the „normal“ S_2 state. For shifts in excitation energies the number of roots to solve is not important. If not stated otherwise, the SORCI calculation solves for 3 states in all calculations in this work.

Table A-7: SORCI(nat19) excitation energies ΔE and weighting factors of the main configurations ($\geq$ 10%) for WT-Rh calculated with the QM1 region with and without external charges solved to 3, 4, and 5 states (order is for similar states and not by energy) (eV).

	without external charges			with all external charges		
	3states	4states	5states	3states	4states	5states[a]
S_0						
22222200	0.83	0.82	0.82	0.80	0.82	0.80
S_1	**1.98**	**1.95**	**1.98**	**2.52**	**2.53**	**2.50**
22222110	0.78	0.76	0.76	0.77	0.78	0.76
S_2	–	**2.87**	**2.94**[b]	**2.70**	**2.71**	**2.62**
22222020	–	0.38	0.22	0.49	0.49	0.42
22222101	–	0.12		0.10		
22212210	–	0.12	0.33			
22221210	–	0.16	0.10	0.18	0.10	0.24
S_3	**2.54**	**2.54**	**2.60**	–	**3.28**	**3.77**
22222020	0.29	0.15	0.19	–		
22221210	0.45	0.59	0.56	–	0.57	0.35
S_4	–	–	**3.72**	–	–	**3.37**
22222020	–	–	0.15	–	–	
22212210	–	–	0.32	–	–	0.46
22221210	–	–	0.11	–	–	0.12
22221120	–	–		–	–	0.13

[a] The SORCI calculation with all external charges solved for 5 states had convergence problems.
[b] S_2 state is strongly changed.

Table A-8: SORCI(nat19) excitation energies ΔE and weighting factors of the main configurations ($\geq$ 10%) calculated with the QM4 region with and without external charges solved to 3 and 4 states (order is for similar states and not by energy) (eV).

a) WT-Rhodopsin

	without external charges		wit all external charges	
	3states	4states	3states	4states
S_0				
22222200	0.79	0.81	0.80	0.82
S_1	**2.76**	**2.69**	**2.53**	**2.55**
22222110	0.75	0.77	0.76	0.77
S_2	**2.73**	**2.69**	**2.70**	**2.66**
22222020	0.45	0.45	0.46	0.45
22222101	0.11	0.10	0.11	0.11
22221210	0.19	0.12	0.19	0.11
S_3	–	**3.31**	–	**3.28**
22221210	–	0.46	–	0.50

b) 30fold mutant (in brackets are the shift in excitation energy relative to WT-Rh)

	without external charges		wit all external charges	
	3states	4states	3states	4states
S_0				
22222200	0.80	0.81	0.81	0.81
S_1	**2.86**(+0.11)	**2.82**(+0.13)	**2.91**(+0.38)	**2.91**(+0.37)
22222110	0.76	0.77	0.78	0.78
S_2	**2.78**(+0.04)	**2.74**(+0.06)	**2.82**(+0.12)	**2.77**(+0.11)
22222020	0.42	0.42	0.42	0.41
22222101	0.12	0.11	0.11	0.11
22212210	<0.03	0.11	<0.03	0.11
22221210	0.21	0.12	0.22	0.13
S_3	–	**3.43**(+0.12)	–	**3.28**(+0.28)
22212210	–	0.11	–	0.28
22221210	–	0.38	–	0.16
22221120	–	0.12	–	0.15

Appendix D

Details on Asp83 and Glu122 Protonation States

Already the mutant FTIR measurements of Fahmy et al.[66] and Lüdeke et al.[12] provide strong evidence for protonated Asp83 and Glu122 residues. They see one disappearing difference band for each single mutant D83N and E122Q and both are missing in the double mutant. Assuming that one (or both) of these residues is deprotonated in WT Rh, the disappearing band in the difference spectrum of its mutant can be explained only in terms of coupled protonation states (see Table A-9a). This means that one residue being protonated in WT-Rh must be unprotonated in a mutant, and one residue being unprotonated in WT-Rh is protonated in a mutant. Additionally these two residues have to have the same vibrational frequency (1767 cm^{-1}). This is not supported by the IR calculations (see Section 4.3.3).

With this argumentation the assignment is very clear, but it checked if a changed assignment would change the results and conclusions. The effect of the protonation states of Asp83 and Glu122 on the spectroscopic features of the chromophore are tested. UV–vis calculations of models with Asp83 and/or Glu122 being unprotonated are performed. The excitation energies with deprotonated Asp83 show very small shifts (Table A-10). This is not surprising, as Asp83 is located 13 Å away from the chromophore and screened by several water molecules (orientational polarization). Deprotonation of Glu122 results in a very large red shift. This can be explained by the small distance to the β-ionone ring. Due to this large shift, the possibility of Glu122 being unprotonated can be excluded because of the large deviation in the excitation energy compared to the experimental value of WT Rh (2.47 eV[57]).

Further experimental evidence for a neutral Glu122 comes from the absorption measurement of the E112L mutant by Lin et al.[117]($\Delta\Delta E_{S_1 \leftarrow S_0} = +0.02\,\mathrm{eV}$). The shifts in excitation energy of this mutant are given in Table A-11. For the models with neutral Glu122, the shifts are small, in agreement with experiment. Mutation shifts of more than 0.6 eV arise when assuming a deprotonated Glu122,

Table A-9: Possible coupling of protonation states of residues Asp83, Glu122, and Glu181. p = protonated, u = unprotonated, (–) = mutated.

(a) With coupling. Asp83 is assumed to be unprotonated. The table is similar for a unprotonated Glu122.

	WT	D83N	E122Q	E181Q	D83N+E122Q
Glu181	p	u	u or p	–	u
Asp83	u	–	p or u	p	–
Glu122	p	p	–	p	–

(b) Without coupling.

	WT	D83N	E122Q	E181Q	D83N+E122Q
Glu181	u	u	u	–	u
Asp83	p	–	p	p	–
Glu122	p	p	–	p	–

regardless of the protonation state of Glu181.

These results clearly show that Glu122 must be protonated, irrespective of the protonation states of Glu181 and Asp83, whereas the protonation state of Asp83 has little influence on the binding site. Taking Asp83 deprotonated would not change the conclusions taken from the UV–Vis and Raman calculation because the excitation energies and also the BLA is very similar (Table A-12 and A-13).

Also NMR chemical shifts are hardly affected by Asp83 deprotonation (Table A-14). The differences are essentially the same for both models and are largest for C5 and C6 ($+2.1$ and -1.4 ppm, respectively). This insignificantly increases the deviation from experiment on these atoms.

Table A-10: Shifts in excitation energy $\Delta\Delta E_{S_1 \leftarrow S_0}$ of models due to deprotonation of Asp83 and/or Glu122 (eV).

model	83+122-prot	83-deprot	122-deprot	83+122-deprot
PROT1	2.43	-0.01	-0.77	-0.80
UNPROT1	2.54	$+0.01$	-0.64	-0.71
GLN1	2.47	$+0.03$	-0.76	-0.78

Table A-11: Shifts in excitation energy $\Delta\Delta E_{S_1 \leftarrow S_0}$ from models with different protonation states of Glu122 to the E122L mutant (SORCI, QM4, eV). The experimental shift between WT-Rh and the E122L mutant is $+0.02$ eV.[117]

model	122-prot	122-deprot
UNPROT1	$+0.01$	$+0.65$
PROT1	-0.02	$+0.67$

Table A-12: Excitation energy $\Delta E_{S_1 \leftarrow S_0}$ of models with different protonation states of Asp83 (SORCI, QM4, eV). The experimental shift between WT Rh and the E181Q mutant is $+0.03$ eV.[57]

model	Asp83-prot	shift to GLN1	Asp83-deprot	shift to GLN1
PROT1	2.43	-0.04	2.42	-0.08
UNPROT1	2.54	$+0.07$	2.55	$+0.05$
GLN1	2.47	–	2.50	–

Table A-13: Bond length alternation (BLA, Å) of models with different protonation states of Asp83.

model	83-prot	83-deprot
PROT1	0.063	0.063
UNPROT1	0.068	0.067
GLN1	0.065	0.065

Table A-14: Difference in ^{13}C NMR chemical shifts due to deprotonation of Asp83 (ppm).

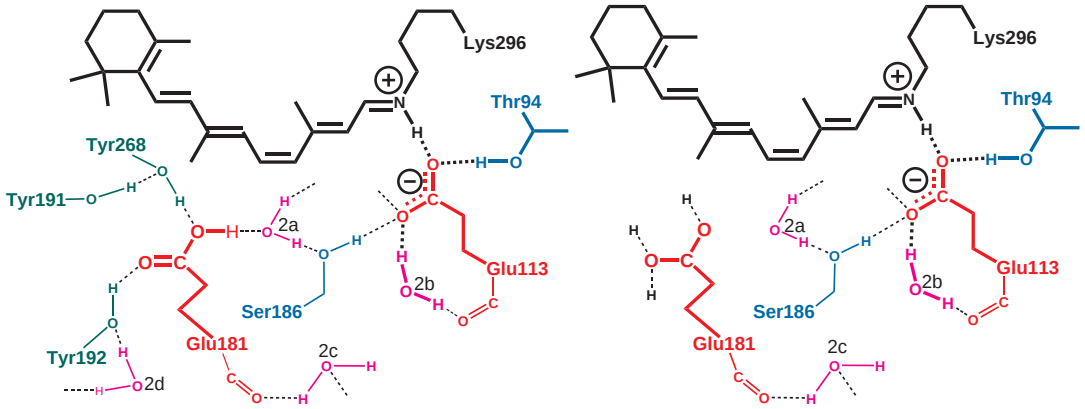
^{13}C	UNPROT1	PROT1	^{13}C	UNPROT1	PROT1
1	-0.1	-0.1	11	0	-0.3
2	0.1	0.2	12	-0.2	0
3	0	0.1	13	0.2	0.1
4	-0.3	-0.3	14	-0.4	0
5	2.4	2.1	15	-0.2	0
6	-1.4	-1.4	16	0.3	0.2
7	0	-0.6	17	0.1	0.1
8	-0.1	0	18	0	0
9	-0.4	-0.8	19	-0.1	-0.1
10	0.1	0.3	20	0.1	0.2

Appendix E

WT-Rh Models

E.1 WT-Rh Models with Protonated Glu181

For the structural model with protonated Glu181, all four possible hydrogen positions at Glu181 are considered. Further, two different orientations of Ser186 (X-ray and rotated orientation) are considered, resulting in eight structural models of the active site, which are shown in Figures A-7 and A-8. They are described with their hydrogen-bonding pattern in Table A-15. For these eight models, QM/MM geometry optimizations with QM4 region are performed. The model with the lowest energy, PROT1, forms an extended HBN ranging from Glu181 to Glu113, as proposed by Okada et al.[52] It is 1.5 kcal/mol more stable than PROT1b. The other models are much higher in energy due to less favorable hydrogen-bonding patterns, their relative energies are shown in Table A-16.



(a) PROT1 **(b)** PROT2–PROT4

Figure A-7: WT-Rh models with protonated Glu181 and Ser186 in rotated orientation.

Additionally the effect of the QM region on the energy difference between mod-

els PROT1 and PROT1b is investigated. When including the Ser186 side chain, Wat2c, and Wat2a in the QM region (QM5_186, QM6_186,2c, QM7_186,2c,2a) the energy difference increases to 4.3, 4.6, and 4.5 kcal/mol, respectively.

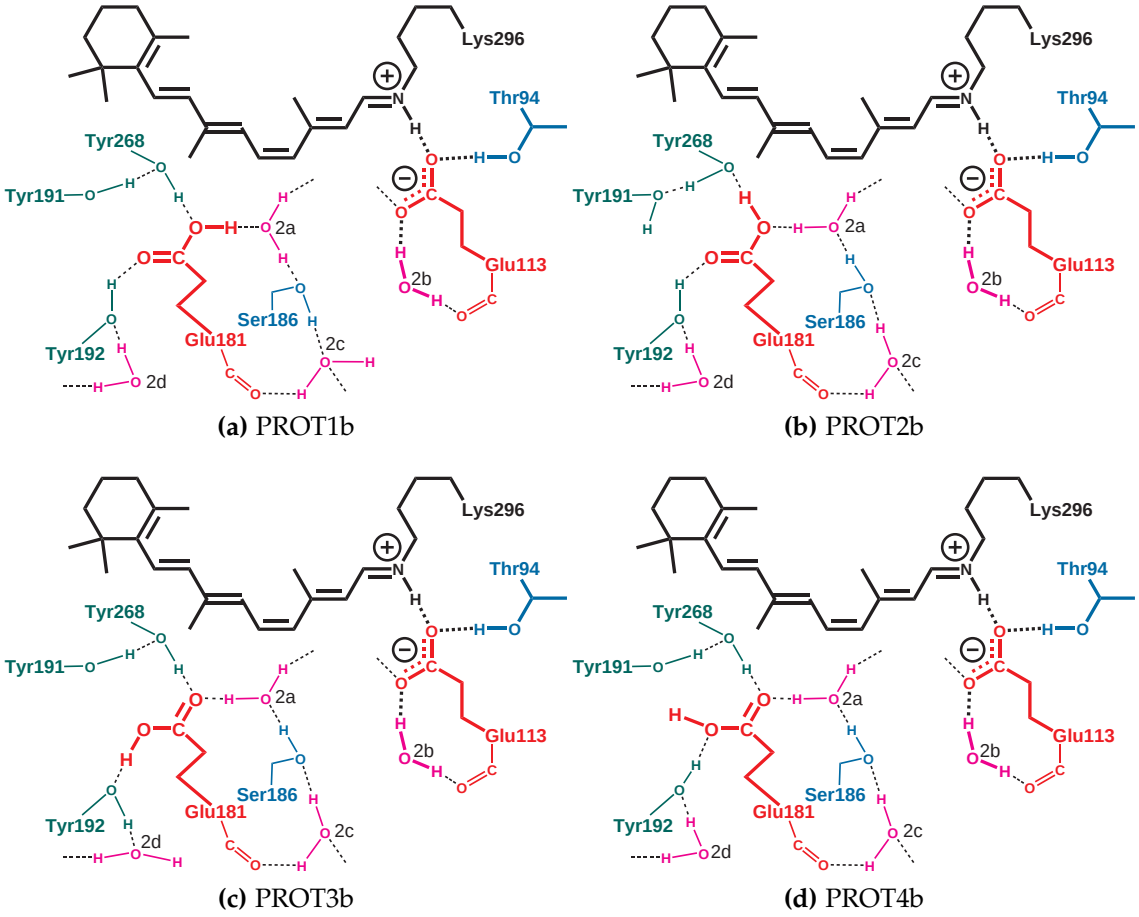
Table A-15: Nomenclature of models with protonated Glu181 and their hydrogen-bonding pattern. [a]

Glu181 hydrogen is	rotated orientation of Ser186	X-ray orientation of Ser186
at $O_{\varepsilon 1}$	**PROT1**	**PROT1b**
and hydrogen-binds	(181-2a-186-113)	(181-2a-186-2c)
to Wat2a	(191-268-181)(2d-192-181)	(191-268-181)(2d-192-181)
at $O_{\varepsilon 1}$	**PROT2**	**PROT2b**
and hydrogen-binds	(2a-186-113)	(186-2a-181)
to Tyr268	(181-268-191*)(2d-192-181)	(181-268-191*)(2d-192-181)
at $O_{\varepsilon 2}$	**PROT3**	**PROT3b**
and hydrogen-binds	(2a-186-113)	(186-2a-181)
to Tyr192	(191-268-181)(181-192-2d*)	(191-268-181)(181-192-2d*)
at $O_{\varepsilon 2}$	**PROT4**	**PROT4b**
and points	(2a-186-113)	(186-2a-181)
away from Tyr192	(191-268-181)(192-181*)	(191-268-181)(192-181*)

[a] The hydrogen-bonding patterns are described in parentheses: E.g., "(191-268-181)" refers to a hydrogen bonds from Tyr191 to Tyr268 and from Tyr268 to Glu181. Residues with dangling OH bonds are marked with "*". "2d" refers to water Wat2d (close to Tyr192).

Table A-16: Relative energies of models with protonated Glu181 (kcal/mol).

Ser186 orientation	PROT1	PROT2	PROT3	PROT4
rotated	0.0	4.9	8.1	6.4
X-ray (PROT*n*b)	**1.5**	10.5	11.5	8.9



(a) PROT1b **(b)** PROT2b

(c) PROT3b **(d)** PROT4b

Figure A-8: Models with protonated Glu181 and Ser186 in X-ray orientation.

E.2 WT-Rh Models with Unprotonated Glu181

Three different models with unprotonated Glu181 were set up:

- UNPROT1 with Ser186 in X-ray orientation (as in PROT1b)
- UNPROT2 with Ser186 in rotated orientation having a hydrogen bond to $O_{\varepsilon 1}$ on Glu113 (as in PROT1)
- UNPROT3 with Ser186 in a similarly rotated orientation having a hydrogen bond to the other oxygen atom, ($O_{\varepsilon 2}$) which is also the hydrogen-bonding partner of the PSB.

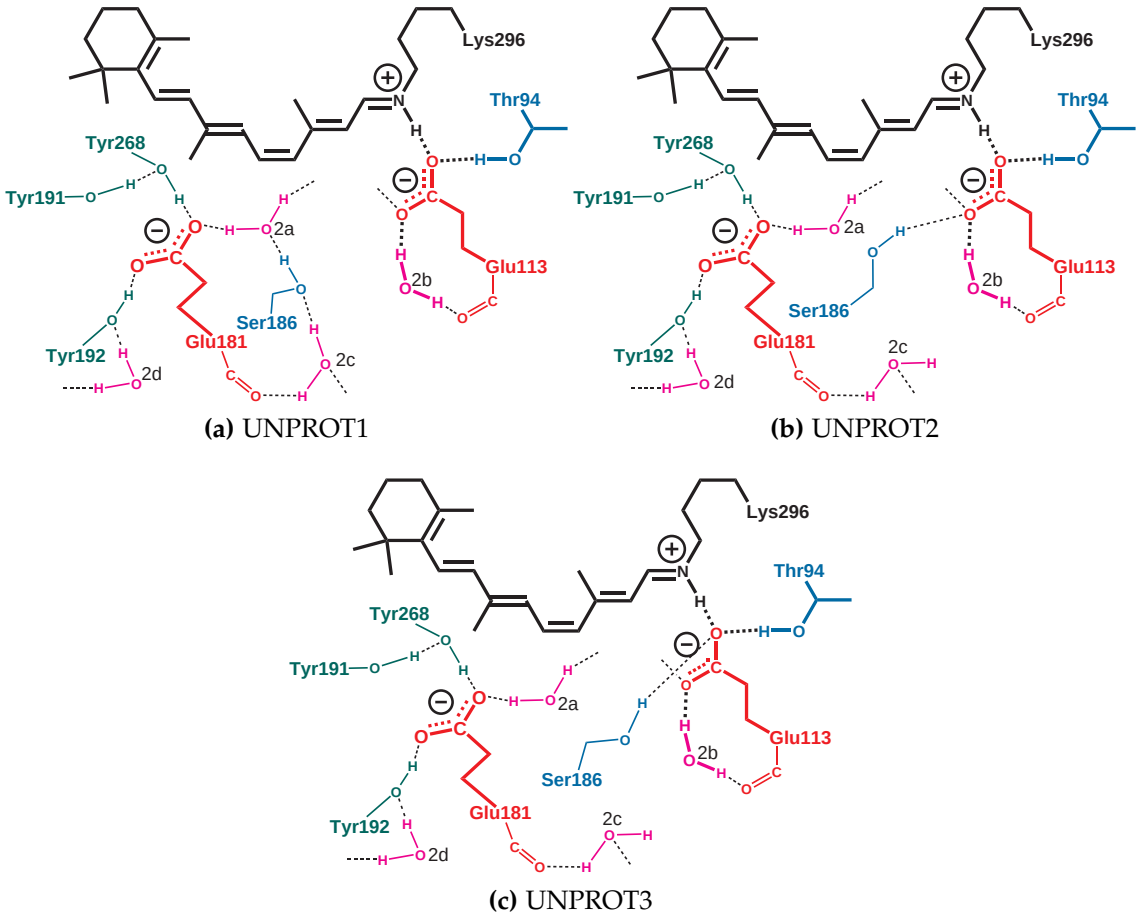
To obtain comparable energies for these models after geometry optimization, accidental configurational differences were avoided by using UNPROT1 as a template and building the other two by modifying merely the HBN. No conformational changes were observed in MD simulations of all tree models. In result, UNPROT2 and UNPROT3 are 10 and 11 kcal/mol higher in energy than

UNPROT1, respectively. In both cases, some oxygen and hydrogen atoms remain without hydrogen-bonding partner. In UNPROT1, a better overall connectivity is achieved, only Glu113 has one hydrogen bond less.

Table A-17: Energies of models with unprotonated Glu181 relative to UNPROT1 (kcal/mol).

	UNPROT2	UNPROT3
QM1	+9.89	+13.11
QM4	+9.95	+10.97

In the MDs of UNPROT1, a translocation of the Tyr191 side chain down to Glu181 is observed when using a small QM region (QM1). This translocation disappears in the larger QM regions (QM4 and QM5_181). The relative energy difference of the minima with and without relocated Tyr191 are −1.3, +0.05, and +1.44 kcal/mol for QM1, QM4, and QM5_181, respectively. This problem demonstrates the limitations of the fixed MM charge models and the importance of the charge exchange from the chromophore to the counterion Glu113.



(a) UNPROT1 **(b)** UNPROT2

(c) UNPROT3

Figure A-9: WT-Rh models with unprotonated Glu181.

Appendix F

Acknowledgements

This work was produced in about five years in the group of Prof. Marcus Elstner at the Institute of Physical an Theoretical Chemistry at TU Braunschweig. I thank him for taking me into his goup although I had no big knowledge on theoretical chemistry at this time. He gave me the topic of color tuning in rhodopsin that was sometimes frustrating because of several new problems appearing again and again. But with his help and assurance that I do good work I never lost motivation completely and now have finished the work in a very good | good | satisfying | fair way.

Starting was hard because of the newly built group for Theoretical Chemistry. Furniture and computer had to be bought and set up. I had to learn several programmes: CHARMM, gnuplot, LINUX, LaTeX, OM2, ORCA, turbomole, VMD (in alphabetic order). Collegues from the group in Paderborn helped me in this. Prasad Phatak, who came to Braunschweig after some months, is an expert in most of these programs and I often needed his help in the first years. Michael Hoffmann teached me how to do CHARMM and OM2 calculations and helped me in doing the mutations for Chapter 5, but then the Problem of Glu181 came up. This work was done together with Marius Wanko, who teached me how to rotate orbitals and to perform polarization calculations with his self-written scripts.

During the whole time several students wrote their "Studienarbeit" in the Theoretical Chemistry group and many of them were guided by me. With the numerous contacts to other institutes I could give them interesting topics to do some calculations on. Most of the totics came from the IHF ("Institut für Hochfrequenztechnik") at TU Braunschweig. Marc Debeaux, Andreas Freund, Daniel Schneidenbach, and Mike Zöllner together with their boss Hans-Hermann Johannes gave me nice molecules for calculations. Iridium complexes for organic light-emitting diodes (OLEDs) and thiophene-nickel(II)porphyrin hybrid molecules. With the latter ones I produced a nice publication with Mike Zöllner.[140] Kai Welke wrote his diploma thesis with my guidance on a similar topic (color tuning in channelrhodopsins). He now started his PhD work and we have a very

interesting collaboration on his and my work although he is in Karlsruhe since some time.

Apart from the research work I had a very nice time in the office. Collegues from the theoretical chemistry group of Marcus Elstner and from the biophysical chemistry group of Jomo Walla made life very nice. Doing pull-ups in the corridor, drinking cappuccino, playing games, and especially the refreshingly trivial talks ("erfrischend belanglosen Diskussionen"[141]) made life at work very pleasant. All names have to be given. In alpabetic order, to not build a hierarchy: Albrecht Goez, Axel Wehling, Carolin König, Christoph Peter Holleboom, Christoph Pilka, Claudia Quentmeier, Johannes Neugebauer, Julia Hoffmann, Kai Welke, Larissa-Nathalie Bieritz, Maike Clemens, Marcus Elstner, Mario Wolter, Michael Gaus, Nour Hafi, Paul Benjamin Woiczikowski, Pen-Nan Liao, Peter Jomo Walla, Prasad Phatak, Stefan Bode, Tino Wolter, Tomáš Kubař, Viktor Werwein, Waldemar Hujo.

Special thanks for proofreading and giving hints for this thesis go to Christoph, Larissa, Kai, Marius, Marco, and Michael. Life apart from the office was short because of the nice collegues. And this short time was filled with the genial work in the "Fachschaft", another group of very good friends being too large to give all names.

Without my family this work would not have been finished. They gave me hope and love and reminded me how proud my grandparents ("Oma und Opa") would have been and will be to have a doctor as grandson.